alpha = 0.05

1 run

training 1 epoch...

44.51965403556824 seconds used......

1 training finished! average train loss: 0.6509972429937787

total clf loss: 114.45371727645397 total adv loss: 62.66376283019781

evaluating...

Confusion matrix, without normalization

[[ 189 78 27]

[ 342 3354 125]

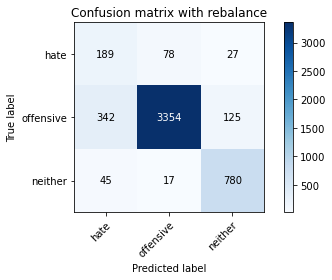
[ 45 17 780]]

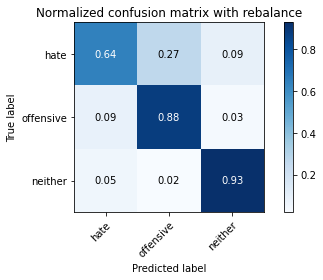
Normalized confusion matrix

[[0.64285714 0.26530612 0.09183673]

[0.08950537 0.87778069 0.03271395]

[0.05344418 0.02019002 0.9263658 ]]





precision recall f1-score support

hate 0.33 0.64 0.43 294

offensive 0.97 0.88 0.92 3821

neither 0.84 0.93 0.88 842

accuracy 0.87 4957

macro avg 0.71 0.82 0.75 4957

weighted avg 0.91 0.87 0.89 4957

average eval\_loss: 0.36756978159950626, accuracy: 0.8721000605204761

training 2 epoch...

44.6242458820343 seconds used......

2 training finished! average train loss: 0.3791057912316936

total clf loss: 67.26284237205982 total adv loss: 48.71504057198763

evaluating...

Confusion matrix, without normalization

[[ 238 39 17]

[ 961 2755 105]

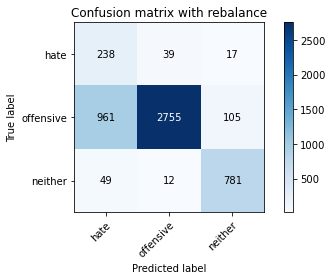
[ 49 12 781]]

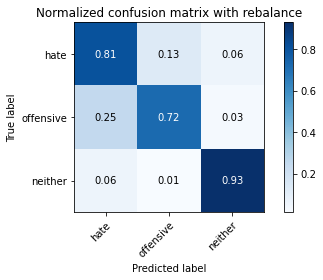
Normalized confusion matrix

[[0.80952381 0.13265306 0.05782313]

[0.25150484 0.72101544 0.02747972]

[0.05819477 0.01425178 0.92755344]]





precision recall f1-score support

hate 0.19 0.81 0.31 294

offensive 0.98 0.72 0.83 3821

neither 0.86 0.93 0.90 842

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: 0.5505038826095481, accuracy: 0.7613475892677022

training 3 epoch...

44.42221426963806 seconds used......

3 training finished! average train loss: 0.2509869848601302

total clf loss: 45.42145583592355 total adv loss: 50.05363220721483

evaluating...

Confusion matrix, without normalization

[[ 212 56 26]

[ 700 2983 138]

[ 23 24 795]]

Normalized confusion matrix

[[0.72108844 0.19047619 0.08843537]

[0.18319812 0.78068568 0.0361162 ]

[0.02731591 0.02850356 0.94418052]]





precision recall f1-score support

hate 0.23 0.72 0.34 294

offensive 0.97 0.78 0.87 3821

neither 0.83 0.94 0.88 842

accuracy 0.80 4957

macro avg 0.68 0.82 0.70 4957

weighted avg 0.90 0.80 0.84 4957

average eval\_loss: 0.49675138233109345, accuracy: 0.8049223320556789

training 4 epoch...

44.47466778755188 seconds used......

4 training finished! average train loss: 0.12478293379365701

total clf loss: 23.84908946417272 total adv loss: 50.22415302693844

evaluating...

Confusion matrix, without normalization

[[ 211 58 25]

[ 590 3116 115]

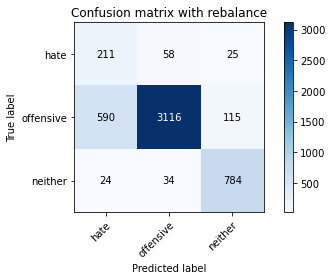
[ 24 34 784]]

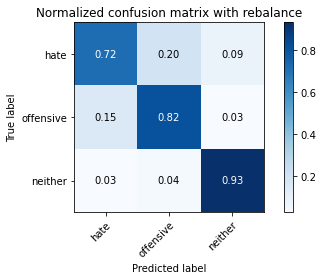
Normalized confusion matrix

[[0.71768707 0.19727891 0.08503401]

[0.15440984 0.81549333 0.03009683]

[0.02850356 0.04038005 0.93111639]]





precision recall f1-score support

hate 0.26 0.72 0.38 294

offensive 0.97 0.82 0.89 3821

neither 0.85 0.93 0.89 842

accuracy 0.83 4957

macro avg 0.69 0.82 0.72 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.48351243836805224, accuracy: 0.8293322574137583

training 5 epoch...

44.50403428077698 seconds used......

5 training finished! average train loss: 0.056158201660067714

total clf loss: 12.208671514876187 total adv loss: 52.11237986758351

evaluating...

Confusion matrix, without normalization

[[ 218 65 11]

[ 706 3015 100]

[ 40 42 760]]

Normalized confusion matrix

[[0.7414966 0.22108844 0.03741497]

[0.18476839 0.78906046 0.02617116]

[0.04750594 0.04988124 0.90261283]]





precision recall f1-score support

hate 0.23 0.74 0.35 294

offensive 0.97 0.79 0.87 3821

neither 0.87 0.90 0.89 842

accuracy 0.81 4957

macro avg 0.69 0.81 0.70 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 0.6757607009382017, accuracy: 0.8055275368166229

training 6 epoch...

44.53798985481262 seconds used......

6 training finished! average train loss: 0.0224421353662136

total clf loss: 6.495501416444313 total adv loss: 53.15792443603277

evaluating...

Confusion matrix, without normalization

[[ 233 45 16]

[ 924 2797 100]

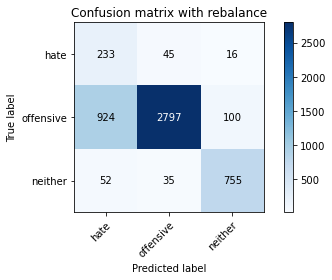
[ 52 35 755]]

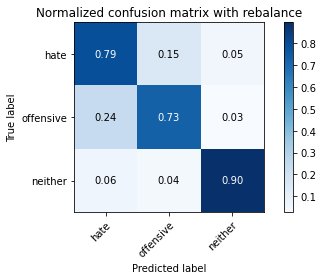
Normalized confusion matrix

[[0.79251701 0.15306122 0.05442177]

[0.24182151 0.73200733 0.02617116]

[0.06175772 0.0415677 0.89667458]]





precision recall f1-score support

hate 0.19 0.79 0.31 294

offensive 0.97 0.73 0.84 3821

neither 0.87 0.90 0.88 842

accuracy 0.76 4957

macro avg 0.68 0.81 0.68 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 1.1346760449270088, accuracy: 0.763566673391164

training 7 epoch...

44.44390082359314 seconds used......

7 training finished! average train loss: 0.03457877966466398

total clf loss: 8.555015685851686 total adv loss: 52.840886786580086

evaluating...

Confusion matrix, without normalization

[[ 251 28 15]

[1163 2562 96]

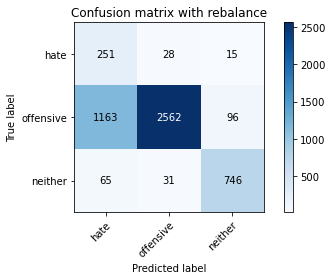
[ 65 31 746]]

Normalized confusion matrix

[[0.8537415 0.0952381 0.05102041]

[0.30437058 0.6705051 0.02512431]

[0.07719715 0.0368171 0.88598575]]





precision recall f1-score support

hate 0.17 0.85 0.28 294

offensive 0.98 0.67 0.80 3821

neither 0.87 0.89 0.88 842

accuracy 0.72 4957

macro avg 0.67 0.80 0.65 4957

weighted avg 0.91 0.72 0.78 4957

average eval\_loss: 1.3541240919186104, accuracy: 0.7179745814000403

training 8 epoch...

44.49120092391968 seconds used......

8 training finished! average train loss: 0.0007244020193945943

total clf loss: 2.7935957480804063 total adv loss: 53.39445888251066

evaluating...

Confusion matrix, without normalization

[[ 220 59 15]

[ 754 2977 90]

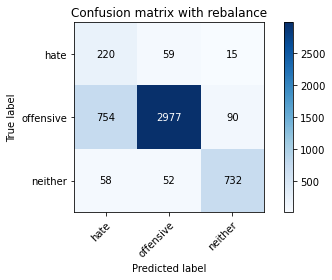
[ 58 52 732]]

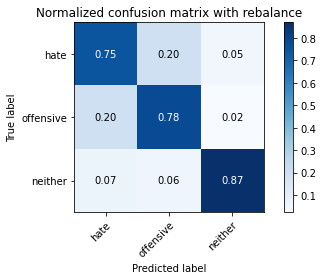
Normalized confusion matrix

[[0.74829932 0.20068027 0.05102041]

[0.19733054 0.77911541 0.02355404]

[0.06888361 0.06175772 0.86935867]]





precision recall f1-score support

hate 0.21 0.75 0.33 294

offensive 0.96 0.78 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 0.90 0.79 0.83 4957

average eval\_loss: 1.0177403623425973, accuracy: 0.7926165019164817

training 9 epoch...

44.42706060409546 seconds used......

9 training finished! average train loss: -0.005882070827357776

total clf loss: 1.7191011400718708 total adv loss: 54.498703978955746

evaluating...

Confusion matrix, without normalization

[[ 233 45 16]

[ 855 2865 101]

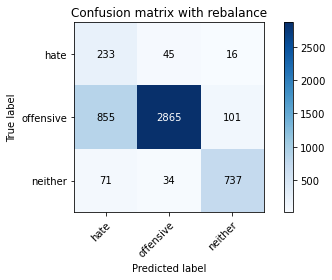
[ 71 34 737]]

Normalized confusion matrix

[[0.79251701 0.15306122 0.05442177]

[0.22376341 0.74980372 0.02643287]

[0.08432304 0.04038005 0.87529691]]





precision recall f1-score support

hate 0.20 0.79 0.32 294

offensive 0.97 0.75 0.85 3821

neither 0.86 0.88 0.87 842

accuracy 0.77 4957

macro avg 0.68 0.81 0.68 4957

weighted avg 0.91 0.77 0.82 4957

average eval\_loss: 1.2690296028966024, accuracy: 0.7736534194068994

training 10 epoch...

44.43128442764282 seconds used......

10 training finished! average train loss: 0.013662352089154704

total clf loss: 5.062617705698358 total adv loss: 54.52710933983326

evaluating...

Confusion matrix, without normalization

[[ 218 57 19]

[ 741 2967 113]

[ 39 43 760]]

Normalized confusion matrix

[[0.7414966 0.19387755 0.06462585]

[0.19392829 0.7764983 0.02957341]

[0.04631829 0.05106888 0.90261283]]





precision recall f1-score support

hate 0.22 0.74 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 0.69 4957

weighted avg 0.90 0.80 0.83 4957

average eval\_loss: 1.0322728498515883, accuracy: 0.795844260641517

2 run

training 1 epoch...

44.531946897506714 seconds used......

1 training finished! average train loss: 0.6608369439317469

total clf loss: 116.10501073300838 total adv loss: 62.03787427395582

evaluating...

Confusion matrix, without normalization

[[ 244 32 18]

[1041 2677 103]

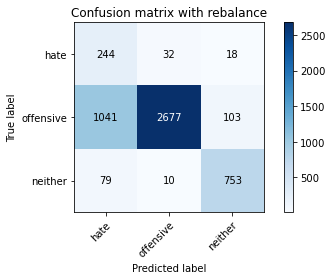
[ 79 10 753]]

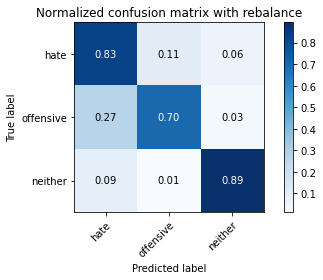
Normalized confusion matrix

[[0.82993197 0.10884354 0.06122449]

[0.27244177 0.70060194 0.02695629]

[0.09382423 0.01187648 0.89429929]]





precision recall f1-score support

hate 0.18 0.83 0.29 294

offensive 0.98 0.70 0.82 3821

neither 0.86 0.89 0.88 842

accuracy 0.74 4957

macro avg 0.67 0.81 0.66 4957

weighted avg 0.92 0.74 0.80 4957

average eval\_loss: 0.6682326011359692, accuracy: 0.7411740972362316

training 2 epoch...

44.529597997665405 seconds used......

2 training finished! average train loss: 0.3866374069574284

total clf loss: 68.54057756066322 total adv loss: 48.511622592806816

evaluating...

Confusion matrix, without normalization

[[ 194 62 38]

[ 482 3170 169]

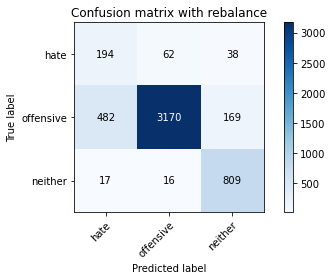
[ 17 16 809]]

Normalized confusion matrix

[[0.65986395 0.21088435 0.1292517 ]

[0.12614499 0.82962575 0.04422926]

[0.02019002 0.01900238 0.9608076 ]]





precision recall f1-score support

hate 0.28 0.66 0.39 294

offensive 0.98 0.83 0.90 3821

neither 0.80 0.96 0.87 842

accuracy 0.84 4957

macro avg 0.68 0.82 0.72 4957

weighted avg 0.90 0.84 0.86 4957

average eval\_loss: 0.4014531510070928, accuracy: 0.8418398224732702

training 3 epoch...

44.57079243659973 seconds used......

3 training finished! average train loss: 0.2620576417062715

total clf loss: 47.23952588438988 total adv loss: 48.55338218063116

evaluating...

Confusion matrix, without normalization

[[ 235 43 16]

[ 875 2856 90]

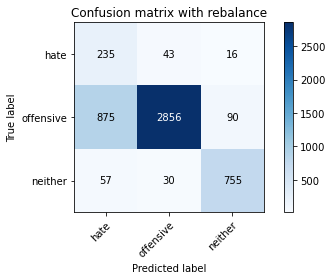
[ 57 30 755]]

Normalized confusion matrix

[[0.79931973 0.1462585 0.05442177]

[0.22899764 0.74744831 0.02355404]

[0.06769596 0.03562945 0.89667458]]





precision recall f1-score support

hate 0.20 0.80 0.32 294

offensive 0.98 0.75 0.85 3821

neither 0.88 0.90 0.89 842

accuracy 0.78 4957

macro avg 0.68 0.81 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.5393814357568419, accuracy: 0.7758725035303611

training 4 epoch...

44.51433038711548 seconds used......

4 training finished! average train loss: 0.15090744544299897

total clf loss: 28.277280014008284 total adv loss: 49.44213683158159

evaluating...

Confusion matrix, without normalization

[[ 239 38 17]

[ 991 2733 97]

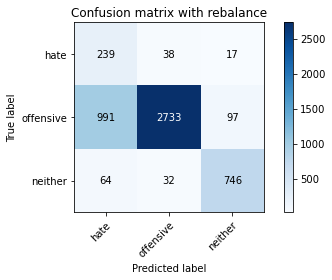
[ 64 32 746]]

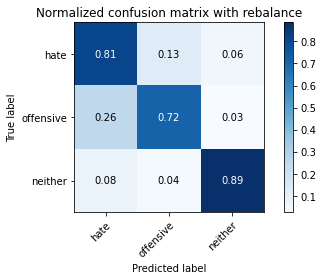
Normalized confusion matrix

[[0.81292517 0.1292517 0.05782313]

[0.25935619 0.71525779 0.02538602]

[0.0760095 0.03800475 0.88598575]]





precision recall f1-score support

hate 0.18 0.81 0.30 294

offensive 0.98 0.72 0.83 3821

neither 0.87 0.89 0.88 842

accuracy 0.75 4957

macro avg 0.68 0.80 0.67 4957

weighted avg 0.91 0.75 0.80 4957

average eval\_loss: 0.6845247135647843, accuracy: 0.7500504337300786

training 5 epoch...

44.51321196556091 seconds used......

5 training finished! average train loss: 0.08318446424642675

total clf loss: 16.722291192039847 total adv loss: 49.95495589077473

evaluating...

Confusion matrix, without normalization

[[ 238 43 13]

[1024 2708 89]

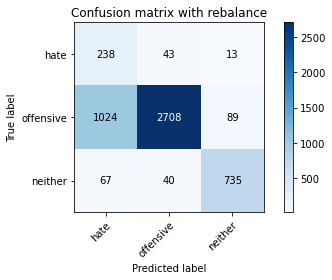
[ 67 40 735]]

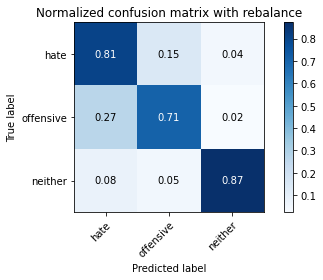
Normalized confusion matrix

[[0.80952381 0.1462585 0.04421769]

[0.26799267 0.708715 0.02329233]

[0.07957245 0.04750594 0.87292162]]





precision recall f1-score support

hate 0.18 0.81 0.29 294

offensive 0.97 0.71 0.82 3821

neither 0.88 0.87 0.88 842

accuracy 0.74 4957

macro avg 0.68 0.80 0.66 4957

weighted avg 0.91 0.74 0.80 4957

average eval\_loss: 0.9011779605531164, accuracy: 0.7425862416784346

training 6 epoch...

44.4339234828949 seconds used......

6 training finished! average train loss: 0.03816243502643635

total clf loss: 9.092583074700087 total adv loss: 51.33613306283951

evaluating...

Confusion matrix, without normalization

[[ 235 37 22]

[1135 2553 133]

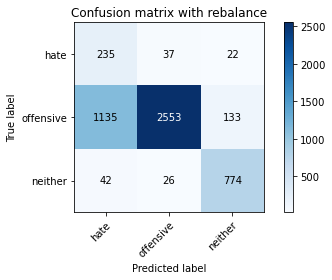
[ 42 26 774]]

Normalized confusion matrix

[[0.79931973 0.12585034 0.07482993]

[0.29704266 0.6681497 0.03480764]

[0.04988124 0.03087886 0.9192399 ]]





precision recall f1-score support

hate 0.17 0.80 0.28 294

offensive 0.98 0.67 0.79 3821

neither 0.83 0.92 0.87 842

accuracy 0.72 4957

macro avg 0.66 0.80 0.65 4957

weighted avg 0.90 0.72 0.78 4957

average eval\_loss: 1.1660704769554637, accuracy: 0.7185797861609845

training 7 epoch...

44.495270013809204 seconds used......

7 training finished! average train loss: 0.02495445345926006

total clf loss: 6.9114618877647445 total adv loss: 52.88500590622425

evaluating...

Confusion matrix, without normalization

[[ 224 52 18]

[ 898 2816 107]

[ 67 42 733]]

Normalized confusion matrix

[[0.76190476 0.17687075 0.06122449]

[0.23501701 0.73697985 0.02800314]

[0.07957245 0.04988124 0.87054632]]





precision recall f1-score support

hate 0.19 0.76 0.30 294

offensive 0.97 0.74 0.84 3821

neither 0.85 0.87 0.86 842

accuracy 0.76 4957

macro avg 0.67 0.79 0.67 4957

weighted avg 0.90 0.76 0.81 4957

average eval\_loss: 0.9728101634180113, accuracy: 0.7611458543473876

training 8 epoch...

44.418437242507935 seconds used......

8 training finished! average train loss: 0.011490221129871948

total clf loss: 4.668072301719803 total adv loss: 54.06488823890686

evaluating...

Confusion matrix, without normalization

[[ 233 46 15]

[ 910 2811 100]

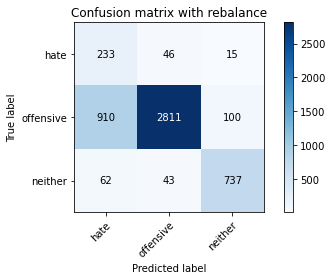
[ 62 43 737]]

Normalized confusion matrix

[[0.79251701 0.15646259 0.05102041]

[0.23815755 0.73567129 0.02617116]

[0.0736342 0.05106888 0.87529691]]





precision recall f1-score support

hate 0.19 0.79 0.31 294

offensive 0.97 0.74 0.84 3821

neither 0.87 0.88 0.87 842

accuracy 0.76 4957

macro avg 0.68 0.80 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 1.1952635606510504, accuracy: 0.7627597337099051

training 9 epoch...

44.4481303691864 seconds used......

9 training finished! average train loss: -0.0013506139699391454

total clf loss: 2.49076635990059 total adv loss: 54.434426479041576

evaluating...

Confusion matrix, without normalization

[[ 216 51 27]

[ 684 3007 130]

[ 26 47 769]]

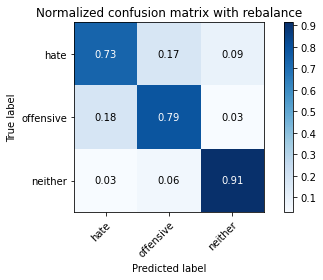
Normalized confusion matrix

[[0.73469388 0.17346939 0.09183673]

[0.17901073 0.78696676 0.03402251]

[0.03087886 0.05581948 0.91330166]]





precision recall f1-score support

hate 0.23 0.73 0.35 294

offensive 0.97 0.79 0.87 3821

neither 0.83 0.91 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: 1.0308191935144995, accuracy: 0.8053258018963082

training 10 epoch...

44.505759954452515 seconds used......

10 training finished! average train loss: 0.0059362240843692715

total clf loss: 3.8603886303608306 total adv loss: 56.90588513761759

evaluating...

Confusion matrix, without normalization

[[ 193 73 28]

[ 452 3205 164]

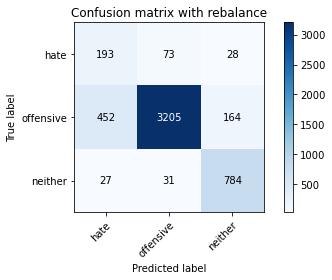
[ 27 31 784]]

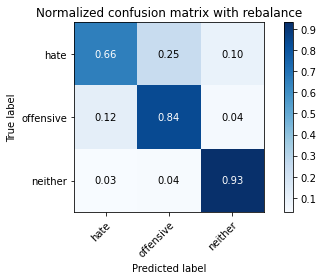
Normalized confusion matrix

[[0.65646259 0.24829932 0.0952381 ]

[0.11829364 0.83878566 0.0429207 ]

[0.03206651 0.0368171 0.93111639]]





precision recall f1-score support

hate 0.29 0.66 0.40 294

offensive 0.97 0.84 0.90 3821

neither 0.80 0.93 0.86 842

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: 0.8115379425538366, accuracy: 0.8436554367561024

3 run

training 1 epoch...

44.48372840881348 seconds used......

1 training finished! average train loss: 0.6326431339247185

total clf loss: 111.1530522853136 total adv loss: 59.421525441110134

evaluating...

Confusion matrix, without normalization

[[ 244 35 15]

[ 910 2791 120]

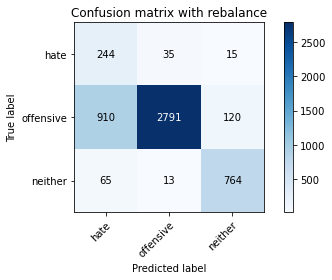
[ 65 13 764]]

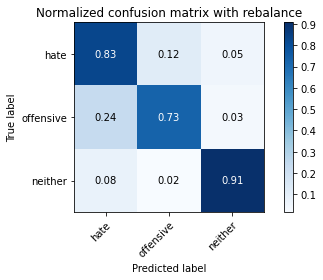
Normalized confusion matrix

[[0.82993197 0.11904762 0.05102041]

[0.23815755 0.73043706 0.03140539]

[0.07719715 0.01543943 0.90736342]]





precision recall f1-score support

hate 0.20 0.83 0.32 294

offensive 0.98 0.73 0.84 3821

neither 0.85 0.91 0.88 842

accuracy 0.77 4957

macro avg 0.68 0.82 0.68 4957

weighted avg 0.91 0.77 0.81 4957

average eval\_loss: 0.5927647370845079, accuracy: 0.7663909622755699

training 2 epoch...

44.52972197532654 seconds used......

2 training finished! average train loss: 0.39167968000759157

total clf loss: 69.51977393776178 total adv loss: 50.85097047686577

evaluating...

Confusion matrix, without normalization

[[ 202 80 12]

[ 397 3317 107]

[ 62 26 754]]

Normalized confusion matrix

[[0.68707483 0.27210884 0.04081633]

[0.1038995 0.86809736 0.02800314]

[0.0736342 0.03087886 0.89548694]]





precision recall f1-score support

hate 0.31 0.69 0.42 294

offensive 0.97 0.87 0.92 3821

neither 0.86 0.90 0.88 842

accuracy 0.86 4957

macro avg 0.71 0.82 0.74 4957

weighted avg 0.91 0.86 0.88 4957

average eval\_loss: 0.3793971540646688, accuracy: 0.8620133145047407

training 3 epoch...

44.410115480422974 seconds used......

3 training finished! average train loss: 0.2832142665154404

total clf loss: 50.94791982322931 total adv loss: 50.36560286581516

evaluating...

Confusion matrix, without normalization

[[ 244 41 9]

[ 909 2852 60]

[ 120 54 668]]

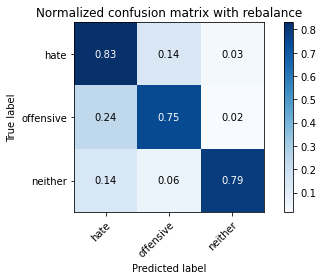
Normalized confusion matrix

[[0.82993197 0.13945578 0.03061224]

[0.23789584 0.74640147 0.0157027 ]

[0.14251781 0.06413302 0.79334917]]





precision recall f1-score support

hate 0.19 0.83 0.31 294

offensive 0.97 0.75 0.84 3821

neither 0.91 0.79 0.85 842

accuracy 0.76 4957

macro avg 0.69 0.79 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 0.5566171298945143, accuracy: 0.7593302400645552

training 4 epoch...

44.446489095687866 seconds used......

4 training finished! average train loss: 0.18201498505359853

total clf loss: 33.668259885162115 total adv loss: 50.87394994497299

evaluating...

Confusion matrix, without normalization

[[ 227 53 14]

[ 691 3055 75]

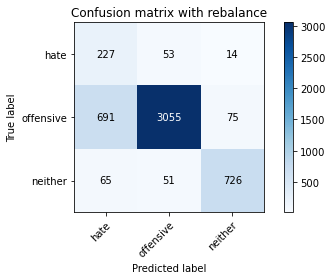
[ 65 51 726]]

Normalized confusion matrix

[[0.77210884 0.18027211 0.04761905]

[0.18084271 0.79952892 0.01962837]

[0.07719715 0.06057007 0.86223278]]





precision recall f1-score support

hate 0.23 0.77 0.36 294

offensive 0.97 0.80 0.88 3821

neither 0.89 0.86 0.88 842

accuracy 0.81 4957

macro avg 0.70 0.81 0.70 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 0.47323610375244773, accuracy: 0.8085535606213435

training 5 epoch...

44.43524193763733 seconds used......

5 training finished! average train loss: 0.09904623726840343

total clf loss: 19.558718462474644 total adv loss: 52.4362366348505

evaluating...

Confusion matrix, without normalization

[[ 235 42 17]

[ 824 2889 108]

[ 45 37 760]]

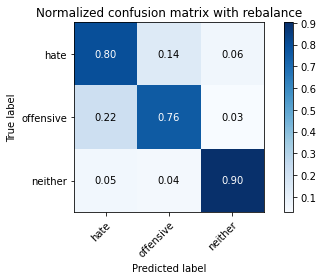
Normalized confusion matrix

[[0.79931973 0.14285714 0.05782313]

[0.21565035 0.75608479 0.02826485]

[0.05344418 0.04394299 0.90261283]]





precision recall f1-score support

hate 0.21 0.80 0.34 294

offensive 0.97 0.76 0.85 3821

neither 0.86 0.90 0.88 842

accuracy 0.78 4957

macro avg 0.68 0.82 0.69 4957

weighted avg 0.91 0.78 0.83 4957

average eval\_loss: 0.6358186883641587, accuracy: 0.78353843050232

training 6 epoch...

44.40371894836426 seconds used......

6 training finished! average train loss: 0.04748373051386392

total clf loss: 10.85389372613281 total adv loss: 54.68351539224386

evaluating...

Confusion matrix, without normalization

[[ 202 71 21]

[ 536 3156 129]

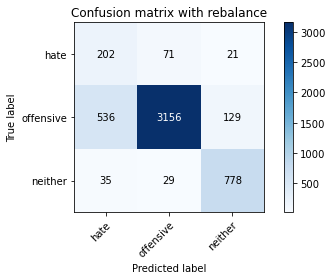
[ 35 29 778]]

Normalized confusion matrix

[[0.68707483 0.2414966 0.07142857]

[0.14027741 0.82596179 0.0337608 ]

[0.0415677 0.03444181 0.9239905 ]]





precision recall f1-score support

hate 0.26 0.69 0.38 294

offensive 0.97 0.83 0.89 3821

neither 0.84 0.92 0.88 842

accuracy 0.83 4957

macro avg 0.69 0.81 0.72 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.6504776579564467, accuracy: 0.834375630421626

training 7 epoch...

44.41199564933777 seconds used......

7 training finished! average train loss: 0.01937153353355825

total clf loss: 5.9859976914012805 total adv loss: 53.46930839121342

evaluating...

Confusion matrix, without normalization

[[ 211 64 19]

[ 588 3126 107]

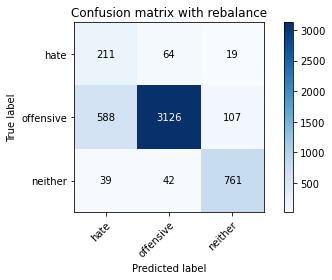
[ 39 42 761]]

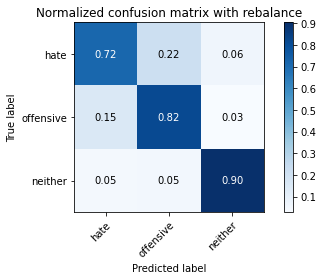
Normalized confusion matrix

[[0.71768707 0.21768707 0.06462585]

[0.15388642 0.81811044 0.02800314]

[0.04631829 0.04988124 0.90380048]]





precision recall f1-score support

hate 0.25 0.72 0.37 294

offensive 0.97 0.82 0.89 3821

neither 0.86 0.90 0.88 842

accuracy 0.83 4957

macro avg 0.69 0.81 0.71 4957

weighted avg 0.91 0.83 0.85 4957

average eval\_loss: 0.692791685850675, accuracy: 0.8267097034496671

training 8 epoch...

44.5695743560791 seconds used......

8 training finished! average train loss: 0.009239979864408573

total clf loss: 4.311639654741157 total adv loss: 54.63206086307764

evaluating...

Confusion matrix, without normalization

[[ 233 42 19]

[ 803 2907 111]

[ 58 32 752]]

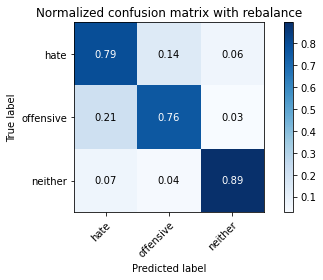
Normalized confusion matrix

[[0.79251701 0.14285714 0.06462585]

[0.21015441 0.7607956 0.02904999]

[0.06888361 0.03800475 0.89311164]]





precision recall f1-score support

hate 0.21 0.79 0.34 294

offensive 0.98 0.76 0.85 3821

neither 0.85 0.89 0.87 842

accuracy 0.79 4957

macro avg 0.68 0.82 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 1.070141960147603, accuracy: 0.7851523098648376

training 9 epoch...

44.42007923126221 seconds used......

9 training finished! average train loss: 0.0021857775894826965

total clf loss: 3.1452501302119344 total adv loss: 55.42964220792055

evaluating...

Confusion matrix, without normalization

[[ 229 49 16]

[ 892 2854 75]

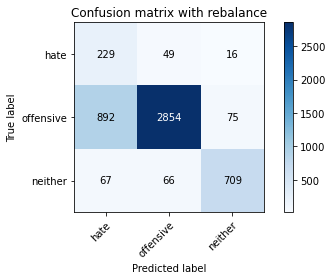
[ 67 66 709]]

Normalized confusion matrix

[[0.77891156 0.16666667 0.05442177]

[0.23344674 0.74692489 0.01962837]

[0.07957245 0.0783848 0.84204276]]





precision recall f1-score support

hate 0.19 0.78 0.31 294

offensive 0.96 0.75 0.84 3821

neither 0.89 0.84 0.86 842

accuracy 0.76 4957

macro avg 0.68 0.79 0.67 4957

weighted avg 0.90 0.76 0.81 4957

average eval\_loss: 1.1379753961274777, accuracy: 0.764978817833367

training 10 epoch...

44.46000099182129 seconds used......

10 training finished! average train loss: 0.011509500743483591

total clf loss: 4.7978017548739444 total adv loss: 56.59354189783335

evaluating...

Confusion matrix, without normalization

[[ 212 63 19]

[ 654 3056 111]

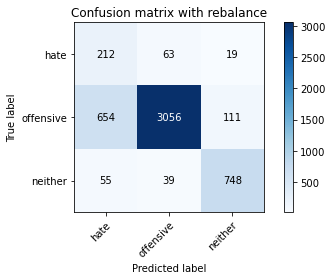
[ 55 39 748]]

Normalized confusion matrix

[[0.72108844 0.21428571 0.06462585]

[0.17115938 0.79979063 0.02904999]

[0.06532067 0.04631829 0.88836105]]





precision recall f1-score support

hate 0.23 0.72 0.35 294

offensive 0.97 0.80 0.88 3821

neither 0.85 0.89 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.80 0.70 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.9225065177808809, accuracy: 0.8101674399838612

4 run

training 1 epoch...

44.49904823303223 seconds used......

1 training finished! average train loss: 0.644020791944356

total clf loss: 113.22210268676281 total adv loss: 61.890952065587044

evaluating...

Confusion matrix, without normalization

[[ 231 43 20]

[ 812 2871 138]

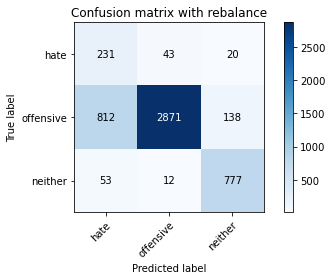
[ 53 12 777]]

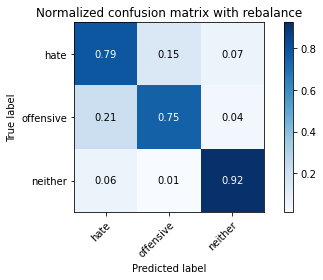
Normalized confusion matrix

[[0.78571429 0.1462585 0.06802721]

[0.21250981 0.75137399 0.0361162 ]

[0.06294537 0.01425178 0.92280285]]





precision recall f1-score support

hate 0.21 0.79 0.33 294

offensive 0.98 0.75 0.85 3821

neither 0.83 0.92 0.87 842

accuracy 0.78 4957

macro avg 0.67 0.82 0.69 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.5286378235586228, accuracy: 0.7825297559007465

training 2 epoch...

44.45308327674866 seconds used......

2 training finished! average train loss: 0.3847049903015644

total clf loss: 68.24892601370811 total adv loss: 49.28745434060693

evaluating...

Confusion matrix, without normalization

[[ 236 46 12]

[ 787 2961 73]

[ 80 34 728]]

Normalized confusion matrix

[[0.80272109 0.15646259 0.04081633]

[0.20596702 0.77492803 0.01910495]

[0.09501188 0.04038005 0.86460808]]





precision recall f1-score support

hate 0.21 0.80 0.34 294

offensive 0.97 0.77 0.86 3821

neither 0.90 0.86 0.88 842

accuracy 0.79 4957

macro avg 0.69 0.81 0.69 4957

weighted avg 0.92 0.79 0.83 4957

average eval\_loss: 0.4711266901163805, accuracy: 0.7918095622352229

training 3 epoch...

44.53744316101074 seconds used......

3 training finished! average train loss: 0.2430712351826025

total clf loss: 44.00647134706378 total adv loss: 48.825803868472576

evaluating...

Confusion matrix, without normalization

[[ 235 45 14]

[ 925 2761 135]

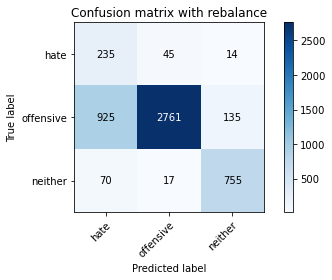
[ 70 17 755]]

Normalized confusion matrix

[[0.79931973 0.15306122 0.04761905]

[0.24208322 0.72258571 0.03533107]

[0.08313539 0.02019002 0.89667458]]





precision recall f1-score support

hate 0.19 0.80 0.31 294

offensive 0.98 0.72 0.83 3821

neither 0.84 0.90 0.86 842

accuracy 0.76 4957

macro avg 0.67 0.81 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 0.6348201291815888, accuracy: 0.756707686100464

training 4 epoch...

44.4782612323761 seconds used......

4 training finished! average train loss: 0.1339384026182761

total clf loss: 25.421049452386796 total adv loss: 50.35165201872587

evaluating...

Confusion matrix, without normalization

[[ 233 48 13]

[ 823 2917 81]

[ 92 38 712]]

Normalized confusion matrix

[[0.79251701 0.16326531 0.04421769]

[0.21538864 0.76341272 0.02119864]

[0.10926366 0.04513064 0.8456057 ]]





precision recall f1-score support

hate 0.20 0.79 0.32 294

offensive 0.97 0.76 0.85 3821

neither 0.88 0.85 0.86 842

accuracy 0.78 4957

macro avg 0.69 0.80 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.7145560254763451, accuracy: 0.7791002622553964

training 5 epoch...

44.45727229118347 seconds used......

5 training finished! average train loss: 0.05434815832332038

total clf loss: 11.874249153770506 total adv loss: 51.614280201494694

evaluating...

Confusion matrix, without normalization

[[ 237 44 13]

[1164 2552 105]

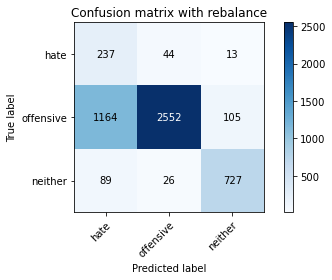
[ 89 26 727]]

Normalized confusion matrix

[[0.80612245 0.14965986 0.04421769]

[0.3046323 0.66788799 0.02747972]

[0.10570071 0.03087886 0.86342043]]





precision recall f1-score support

hate 0.16 0.81 0.27 294

offensive 0.97 0.67 0.79 3821

neither 0.86 0.86 0.86 842

accuracy 0.71 4957

macro avg 0.66 0.78 0.64 4957

weighted avg 0.91 0.71 0.77 4957

average eval\_loss: 1.0911070181717795, accuracy: 0.7092999798265079

training 6 epoch...

44.38974452018738 seconds used......

6 training finished! average train loss: 0.026497774545037954

total clf loss: 7.08203391317511 total adv loss: 51.01828941702843

evaluating...

Confusion matrix, without normalization

[[ 195 80 19]

[ 431 3274 116]

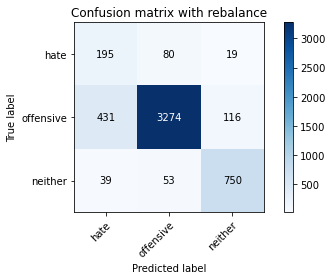
[ 39 53 750]]

Normalized confusion matrix

[[0.66326531 0.27210884 0.06462585]

[0.1127977 0.85684376 0.03035854]

[0.04631829 0.06294537 0.89073634]]





precision recall f1-score support

hate 0.29 0.66 0.41 294

offensive 0.96 0.86 0.91 3821

neither 0.85 0.89 0.87 842

accuracy 0.85 4957

macro avg 0.70 0.80 0.73 4957

weighted avg 0.90 0.85 0.87 4957

average eval\_loss: 0.600349259329021, accuracy: 0.8511196288077466

training 7 epoch...

44.404961824417114 seconds used......

7 training finished! average train loss: 0.01919157002992623

total clf loss: 6.0320040002116 total adv loss: 55.00490888208151

evaluating...

Confusion matrix, without normalization

[[ 226 50 18]

[ 883 2793 145]

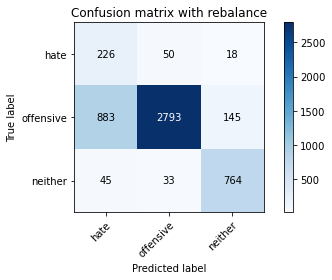
[ 45 33 764]]

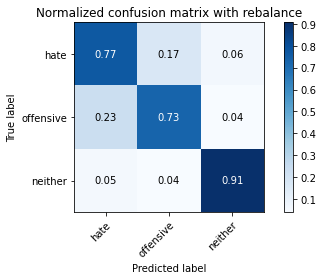
Normalized confusion matrix

[[0.76870748 0.17006803 0.06122449]

[0.23109134 0.73096048 0.03794818]

[0.05344418 0.0391924 0.90736342]]





precision recall f1-score support

hate 0.20 0.77 0.31 294

offensive 0.97 0.73 0.83 3821

neither 0.82 0.91 0.86 842

accuracy 0.76 4957

macro avg 0.66 0.80 0.67 4957

weighted avg 0.90 0.76 0.81 4957

average eval\_loss: 1.1169264016733054, accuracy: 0.7631632035505346

training 8 epoch...

44.411492586135864 seconds used......

8 training finished! average train loss: 0.003602670609123177

total clf loss: 3.2794706408749335 total adv loss: 53.2682780623436

evaluating...

Confusion matrix, without normalization

[[ 228 48 18]

[ 818 2844 159]

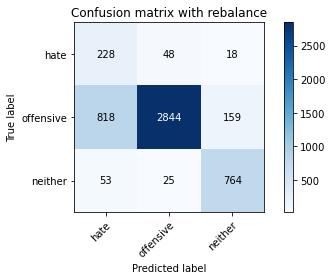
[ 53 25 764]]

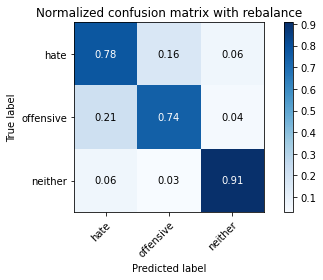
Normalized confusion matrix

[[0.7755102 0.16326531 0.06122449]

[0.21408008 0.74430777 0.04161214]

[0.06294537 0.02969121 0.90736342]]





precision recall f1-score support

hate 0.21 0.78 0.33 294

offensive 0.97 0.74 0.84 3821

neither 0.81 0.91 0.86 842

accuracy 0.77 4957

macro avg 0.66 0.81 0.68 4957

weighted avg 0.90 0.77 0.82 4957

average eval\_loss: 1.2858286429977706, accuracy: 0.7738551543272141

training 9 epoch...

44.525614738464355 seconds used......

9 training finished! average train loss: -0.008470939830413948

total clf loss: 1.339038513979176 total adv loss: 55.75138342380524

evaluating...

Confusion matrix, without normalization

[[ 238 42 14]

[1018 2722 81]

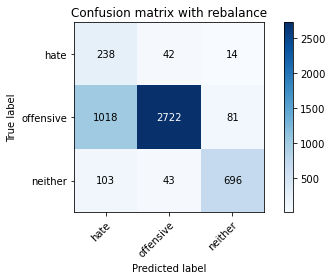
[ 103 43 696]]

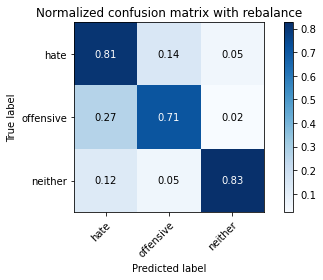
Normalized confusion matrix

[[0.80952381 0.14285714 0.04761905]

[0.2664224 0.71237896 0.02119864]

[0.12232779 0.05106888 0.82660333]]





precision recall f1-score support

hate 0.18 0.81 0.29 294

offensive 0.97 0.71 0.82 3821

neither 0.88 0.83 0.85 842

accuracy 0.74 4957

macro avg 0.67 0.78 0.65 4957

weighted avg 0.91 0.74 0.80 4957

average eval\_loss: 1.487427832423559, accuracy: 0.7375428686705668

training 10 epoch...

44.45447516441345 seconds used......

10 training finished! average train loss: -0.0037346834354974024

total clf loss: 2.173344709270168 total adv loss: 56.23951054364443

evaluating...

Confusion matrix, without normalization

[[ 203 74 17]

[ 564 3150 107]

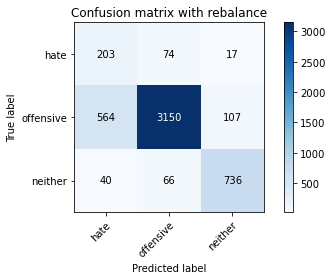
[ 40 66 736]]

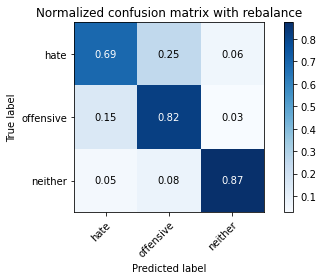
Normalized confusion matrix

[[0.69047619 0.25170068 0.05782313]

[0.14760534 0.82439152 0.02800314]

[0.04750594 0.0783848 0.87410926]]





precision recall f1-score support

hate 0.25 0.69 0.37 294

offensive 0.96 0.82 0.89 3821

neither 0.86 0.87 0.86 842

accuracy 0.82 4957

macro avg 0.69 0.80 0.71 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.936285505833615, accuracy: 0.8248940891668348

5 run

training 1 epoch...

44.39829683303833 seconds used......

1 training finished! average train loss: 0.6315842523560886

total clf loss: 110.97366753220558 total adv loss: 59.455207608640194

evaluating...

Confusion matrix, without normalization

[[ 190 85 19]

[ 436 3297 88]

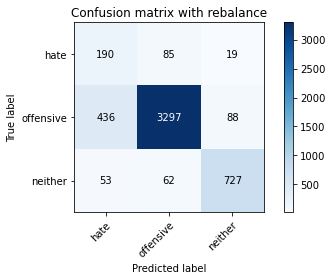
[ 53 62 727]]

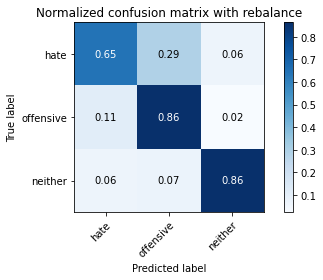
Normalized confusion matrix

[[0.6462585 0.28911565 0.06462585]

[0.11410625 0.86286312 0.02303062]

[0.06294537 0.0736342 0.86342043]]





precision recall f1-score support

hate 0.28 0.65 0.39 294

offensive 0.96 0.86 0.91 3821

neither 0.87 0.86 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: 0.3639142877993084, accuracy: 0.8501109542061731

training 2 epoch...

44.418118476867676 seconds used......

2 training finished! average train loss: 0.384061243145438

total clf loss: 68.27321894466877 total adv loss: 51.974923729896545

evaluating...

Confusion matrix, without normalization

[[ 229 46 19]

[ 795 2870 156]

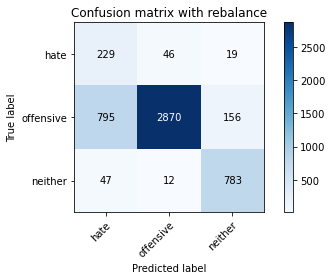
[ 47 12 783]]

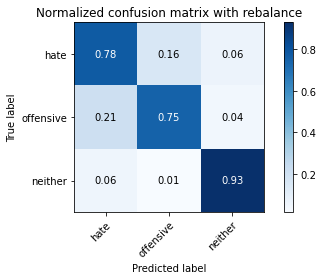
Normalized confusion matrix

[[0.77891156 0.15646259 0.06462585]

[0.20806072 0.75111227 0.04082701]

[0.05581948 0.01425178 0.92992874]]





precision recall f1-score support

hate 0.21 0.78 0.34 294

offensive 0.98 0.75 0.85 3821

neither 0.82 0.93 0.87 842

accuracy 0.78 4957

macro avg 0.67 0.82 0.69 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.5183669959585513, accuracy: 0.7831349606616905

training 3 epoch...

44.40022373199463 seconds used......

3 training finished! average train loss: 0.2584913201978681

total clf loss: 46.63828206062317 total adv loss: 48.7253287602216

evaluating...

Confusion matrix, without normalization

[[ 249 32 13]

[1163 2566 92]

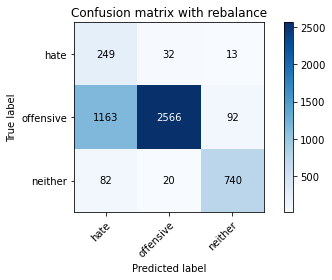
[ 82 20 740]]

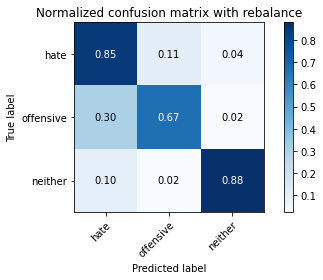
Normalized confusion matrix

[[0.84693878 0.10884354 0.04421769]

[0.30437058 0.67155195 0.02407747]

[0.09738717 0.02375297 0.87885986]]





precision recall f1-score support

hate 0.17 0.85 0.28 294

offensive 0.98 0.67 0.80 3821

neither 0.88 0.88 0.88 842

accuracy 0.72 4957

macro avg 0.67 0.80 0.65 4957

weighted avg 0.91 0.72 0.78 4957

average eval\_loss: 0.6790335892789787, accuracy: 0.7171676417187816

training 4 epoch...

44.397512435913086 seconds used......

4 training finished! average train loss: 0.14028855287932862

total clf loss: 26.49527512397617 total adv loss: 50.11865074932575

evaluating...

Confusion matrix, without normalization

[[ 196 68 30]

[ 488 3062 271]

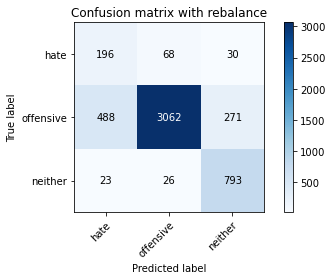
[ 23 26 793]]

Normalized confusion matrix

[[0.66666667 0.23129252 0.10204082]

[0.12771526 0.8013609 0.07092384]

[0.02731591 0.03087886 0.94180523]]





precision recall f1-score support

hate 0.28 0.67 0.39 294

offensive 0.97 0.80 0.88 3821

neither 0.72 0.94 0.82 842

accuracy 0.82 4957

macro avg 0.66 0.80 0.70 4957

weighted avg 0.89 0.82 0.84 4957

average eval\_loss: 0.5271281145452973, accuracy: 0.8172281621948759

training 5 epoch...

44.37486433982849 seconds used......

5 training finished! average train loss: 0.05992987092682406

total clf loss: 12.793399289716035 total adv loss: 50.9078254327178

evaluating...

Confusion matrix, without normalization

[[ 224 47 23]

[ 786 2887 148]

[ 47 24 771]]

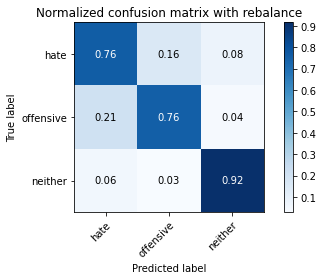
Normalized confusion matrix

[[0.76190476 0.15986395 0.07823129]

[0.20570531 0.75556137 0.03873332]

[0.05581948 0.02850356 0.91567696]]





precision recall f1-score support

hate 0.21 0.76 0.33 294

offensive 0.98 0.76 0.85 3821

neither 0.82 0.92 0.86 842

accuracy 0.78 4957

macro avg 0.67 0.81 0.68 4957

weighted avg 0.90 0.78 0.82 4957

average eval\_loss: 0.7706918919932157, accuracy: 0.7831349606616905

training 6 epoch...

44.45106792449951 seconds used......

6 training finished! average train loss: 0.01576808307657552

total clf loss: 5.263197716674767 total adv loss: 51.337108828127384

evaluating...

Confusion matrix, without normalization

[[ 223 51 20]

[ 736 2973 112]

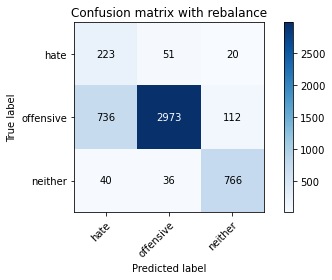
[ 40 36 766]]

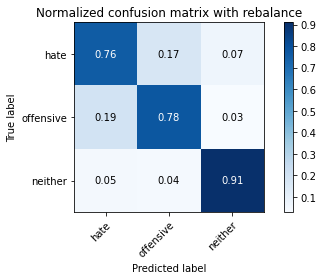
Normalized confusion matrix

[[0.7585034 0.17346939 0.06802721]

[0.19261973 0.77806857 0.0293117 ]

[0.04750594 0.04275534 0.90973872]]





precision recall f1-score support

hate 0.22 0.76 0.34 294

offensive 0.97 0.78 0.86 3821

neither 0.85 0.91 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: 0.927931236903063, accuracy: 0.7992737542868671

training 7 epoch...

44.38409781455994 seconds used......

7 training finished! average train loss: 0.00534689634744274

total clf loss: 3.5736623604316264 total adv loss: 53.18686097860336

evaluating...

Confusion matrix, without normalization

[[ 227 49 18]

[ 809 2906 106]

[ 61 30 751]]

Normalized confusion matrix

[[0.77210884 0.16666667 0.06122449]

[0.21172468 0.76053389 0.02774143]

[0.07244656 0.03562945 0.89192399]]





precision recall f1-score support

hate 0.21 0.77 0.33 294

offensive 0.97 0.76 0.85 3821

neither 0.86 0.89 0.87 842

accuracy 0.78 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.78 0.83 4957

average eval\_loss: 0.9801900799202192, accuracy: 0.78353843050232

training 8 epoch...

44.40628933906555 seconds used......

8 training finished! average train loss: -0.0016239749360284958

total clf loss: 2.3925151027506217 total adv loss: 53.40429547429085

evaluating...

Confusion matrix, without normalization

[[ 233 43 18]

[ 916 2786 119]

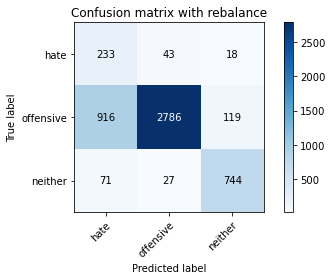
[ 71 27 744]]

Normalized confusion matrix

[[0.79251701 0.1462585 0.06122449]

[0.23972782 0.7291285 0.03114368]

[0.08432304 0.03206651 0.88361045]]





precision recall f1-score support

hate 0.19 0.79 0.31 294

offensive 0.98 0.73 0.83 3821

neither 0.84 0.88 0.86 842

accuracy 0.76 4957

macro avg 0.67 0.80 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 1.262341527170652, accuracy: 0.7591285051442405

training 9 epoch...

44.377859115600586 seconds used......

9 training finished! average train loss: -0.004910963611971391

total clf loss: 1.924915152718313 total adv loss: 55.29379799216986

evaluating...

Confusion matrix, without normalization

[[ 224 59 11]

[ 671 3078 72]

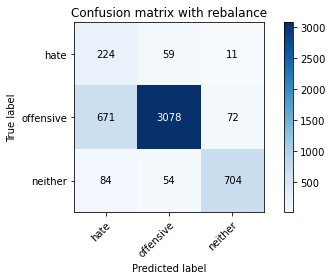
[ 84 54 704]]

Normalized confusion matrix

[[0.76190476 0.20068027 0.03741497]

[0.17560848 0.80554829 0.01884323]

[0.09976247 0.06413302 0.83610451]]





precision recall f1-score support

hate 0.23 0.76 0.35 294

offensive 0.96 0.81 0.88 3821

neither 0.89 0.84 0.86 842

accuracy 0.81 4957

macro avg 0.70 0.80 0.70 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 0.9693001328747824, accuracy: 0.8081500907807141

training 10 epoch...

44.410314321517944 seconds used......

10 training finished! average train loss: -0.0009341783704486681

total clf loss: 2.8037049853010103 total adv loss: 59.26898906379938

evaluating...

Confusion matrix, without normalization

[[ 198 79 17]

[ 586 3144 91]

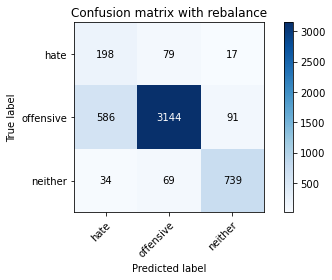
[ 34 69 739]]

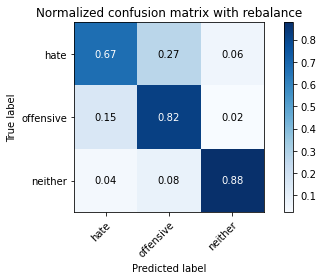
Normalized confusion matrix

[[0.67346939 0.26870748 0.05782313]

[0.15336299 0.82282125 0.02381576]

[0.04038005 0.08194774 0.87767221]]





precision recall f1-score support

hate 0.24 0.67 0.36 294

offensive 0.96 0.82 0.88 3821

neither 0.87 0.88 0.88 842

accuracy 0.82 4957

macro avg 0.69 0.79 0.71 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.877717319141788, accuracy: 0.8232802098043172