hate as no offensive

alpha=2

1 run......

Downloading: 100%

411/411 [00:00<00:00, 10.8kB/s]

Downloading: 100%

263M/263M [00:12<00:00, 23.7MB/s]

training 1 epoch...

843.3872845172882 seconds used......

1 training finished! average train loss: 0.7547125962803201

total clf loss: 282.2593981074169 total adv loss: 233.20774875907227

evaluating...

Confusion matrix, without normalization

[[ 113 154 27]

[ 111 3621 89]

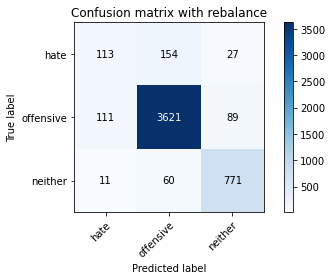
[ 11 60 771]]

Normalized confusion matrix

[[0.38435374 0.52380952 0.09183673]

[0.02904999 0.94765768 0.02329233]

[0.01306413 0.07125891 0.91567696]]





precision recall f1-score support

hate 0.48 0.38 0.43 294

offensive 0.94 0.95 0.95 3821

neither 0.87 0.92 0.89 842

accuracy 0.91 4957

macro avg 0.76 0.75 0.75 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.2434155779379991, accuracy: 0.9088158160177526

training 2 epoch...

842.4943375587463 seconds used......

2 training finished! average train loss: 0.5708151105772553

total clf loss: 208.03745403233916 total adv loss: 179.10556773096323

evaluating...

Confusion matrix, without normalization

[[ 46 220 28]

[ 23 3720 78]

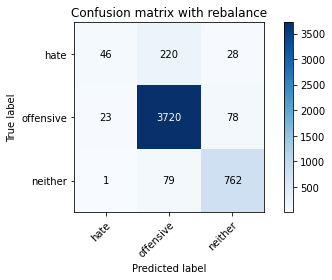
[ 1 79 762]]

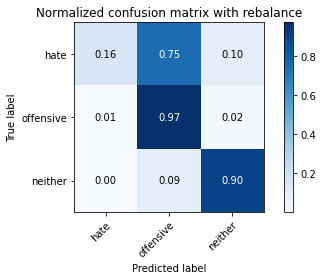
Normalized confusion matrix

[[0.15646259 0.74829932 0.0952381 ]

[0.00601937 0.97356713 0.0204135 ]

[0.00118765 0.09382423 0.90498812]]





precision recall f1-score support

hate 0.66 0.16 0.25 294

offensive 0.93 0.97 0.95 3821

neither 0.88 0.90 0.89 842

accuracy 0.91 4957

macro avg 0.82 0.68 0.70 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.24762710301400792, accuracy: 0.9134557191849909

training 3 epoch...

843.5945763587952 seconds used......

3 training finished! average train loss: 0.4411371397455373

total clf loss: 159.6360343056731 total adv loss: 138.9860043900553

evaluating...

Confusion matrix, without normalization

[[ 89 192 13]

[ 89 3683 49]

[ 26 113 703]]

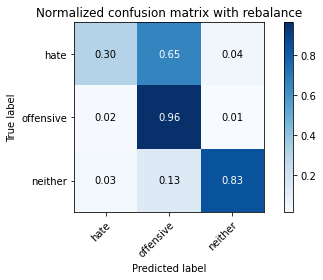
Normalized confusion matrix

[[0.30272109 0.65306122 0.04421769]

[0.02329233 0.9638838 0.01282387]

[0.03087886 0.13420428 0.83491686]]





precision recall f1-score support

hate 0.44 0.30 0.36 294

offensive 0.92 0.96 0.94 3821

neither 0.92 0.83 0.87 842

accuracy 0.90 4957

macro avg 0.76 0.70 0.73 4957

weighted avg 0.89 0.90 0.90 4957

average eval\_loss: 0.2684929111856036, accuracy: 0.9027637684083115

training 4 epoch...

842.2019944190979 seconds used......

4 training finished! average train loss: 0.30201352452915614

total clf loss: 109.7625053304364 total adv loss: 94.91745545557933

evaluating...

Confusion matrix, without normalization

[[ 145 127 22]

[ 211 3519 91]

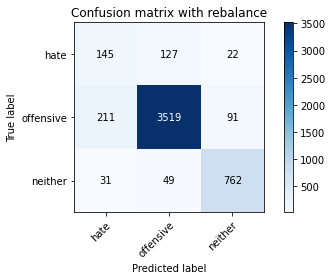
[ 31 49 762]]

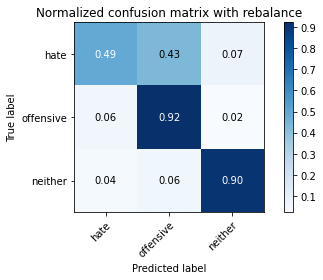
Normalized confusion matrix

[[0.49319728 0.43197279 0.07482993]

[0.05522115 0.9209631 0.02381576]

[0.0368171 0.05819477 0.90498812]]





precision recall f1-score support

hate 0.37 0.49 0.43 294

offensive 0.95 0.92 0.94 3821

neither 0.87 0.90 0.89 842

accuracy 0.89 4957

macro avg 0.73 0.77 0.75 4957

weighted avg 0.90 0.89 0.90 4957

average eval\_loss: 0.3172309100853219, accuracy: 0.8928787573128909

training 5 epoch...

841.9447844028473 seconds used......

5 training finished! average train loss: 0.16979108296589168

total clf loss: 62.59969326169812 total adv loss: 52.91653064016282

evaluating...

Confusion matrix, without normalization

[[ 136 149 9]

[ 226 3558 37]

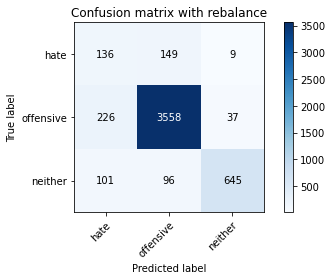
[ 101 96 645]]

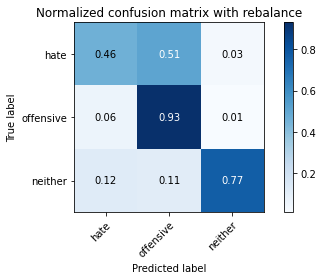
Normalized confusion matrix

[[0.46258503 0.50680272 0.03061224]

[0.05914682 0.93116985 0.00968333]

[0.11995249 0.11401425 0.76603325]]





precision recall f1-score support

hate 0.29 0.46 0.36 294

offensive 0.94 0.93 0.93 3821

neither 0.93 0.77 0.84 842

accuracy 0.88 4957

macro avg 0.72 0.72 0.71 4957

weighted avg 0.90 0.88 0.88 4957

average eval\_loss: 0.42032677546751895, accuracy: 0.8753278192455114

2 run......

training 1 epoch...

842.0925602912903 seconds used......

1 training finished! average train loss: 0.7477921206511617

total clf loss: 282.79639907646924 total adv loss: 229.50669168122113

evaluating...

Confusion matrix, without normalization

[[ 115 149 30]

[ 104 3634 83]

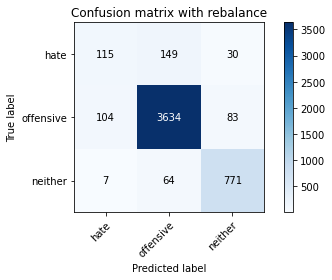
[ 7 64 771]]

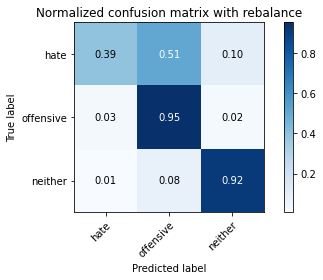
Normalized confusion matrix

[[0.39115646 0.50680272 0.10204082]

[0.02721801 0.95105993 0.02172206]

[0.00831354 0.0760095 0.91567696]]





precision recall f1-score support

hate 0.51 0.39 0.44 294

offensive 0.94 0.95 0.95 3821

neither 0.87 0.92 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.75 0.76 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.24599366572745626, accuracy: 0.9118418398224732

training 2 epoch...

842.7504422664642 seconds used......

2 training finished! average train loss: 0.5713305539308837

total clf loss: 209.19851484615356 total adv loss: 178.78069742396474

evaluating...

Confusion matrix, without normalization

[[ 68 206 20]

[ 41 3708 72]

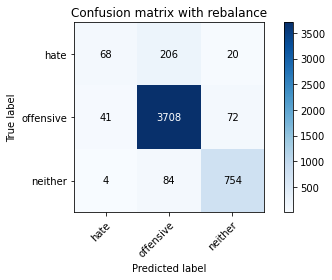
[ 4 84 754]]

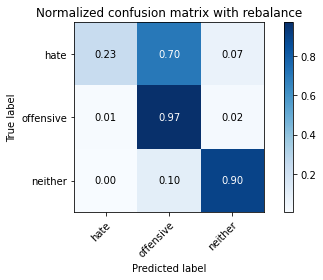
Normalized confusion matrix

[[0.23129252 0.70068027 0.06802721]

[0.01073018 0.97042659 0.01884323]

[0.00475059 0.09976247 0.89548694]]





precision recall f1-score support

hate 0.60 0.23 0.33 294

offensive 0.93 0.97 0.95 3821

neither 0.89 0.90 0.89 842

accuracy 0.91 4957

macro avg 0.81 0.70 0.73 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.2396122682614312, accuracy: 0.9138591890256204

training 3 epoch...

843.0208282470703 seconds used......

3 training finished! average train loss: 0.4450347858308364

total clf loss: 160.87998630385846 total adv loss: 140.2972612627782

evaluating...

Confusion matrix, without normalization

[[ 76 203 15]

[ 52 3718 51]

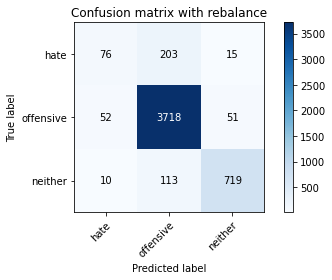
[ 10 113 719]]

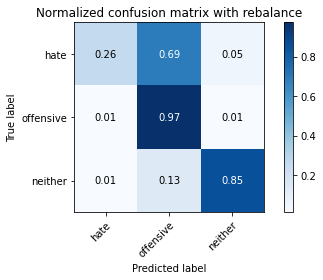
Normalized confusion matrix

[[0.2585034 0.69047619 0.05102041]

[0.013609 0.97304371 0.01334729]

[0.01187648 0.13420428 0.85391924]]





precision recall f1-score support

hate 0.55 0.26 0.35 294

offensive 0.92 0.97 0.95 3821

neither 0.92 0.85 0.88 842

accuracy 0.91 4957

macro avg 0.80 0.70 0.73 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.2781574536347762, accuracy: 0.9104296953802703

training 4 epoch...

842.9330422878265 seconds used......

4 training finished! average train loss: 0.30174528958178276

total clf loss: 109.7754814713262 total adv loss: 94.77792327548377

evaluating...

Confusion matrix, without normalization

[[ 104 174 16]

[ 108 3652 61]

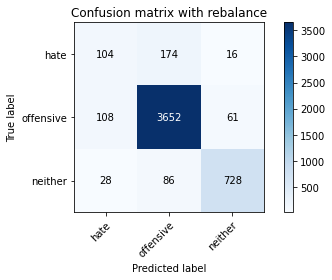
[ 28 86 728]]

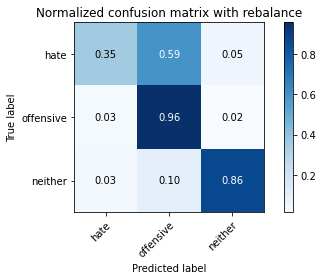
Normalized confusion matrix

[[0.3537415 0.59183673 0.05442177]

[0.02826485 0.95577074 0.01596441]

[0.03325416 0.10213777 0.86460808]]





precision recall f1-score support

hate 0.43 0.35 0.39 294

offensive 0.93 0.96 0.94 3821

neither 0.90 0.86 0.88 842

accuracy 0.90 4957

macro avg 0.76 0.72 0.74 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.29746916856158584, accuracy: 0.9045793826911438

training 5 epoch...

842.643152475357 seconds used......

5 training finished! average train loss: 0.17479430517672764

total clf loss: 64.44896099180914 total adv loss: 54.47349484446022

evaluating...

Confusion matrix, without normalization

[[ 102 172 20]

[ 99 3652 70]

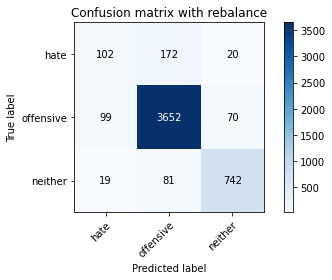
[ 19 81 742]]

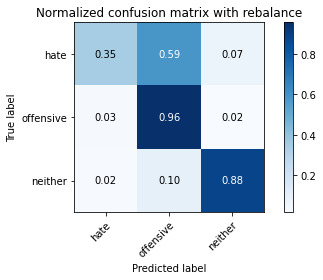
Normalized confusion matrix

[[0.34693878 0.58503401 0.06802721]

[0.02590945 0.95577074 0.01831981]

[0.02256532 0.09619952 0.88123515]]





precision recall f1-score support

hate 0.46 0.35 0.40 294

offensive 0.94 0.96 0.95 3821

neither 0.89 0.88 0.89 842

accuracy 0.91 4957

macro avg 0.76 0.73 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.4056741467234866, accuracy: 0.9070002017349204

3 run......

training 1 epoch...

843.0863089561462 seconds used......

1 training finished! average train loss: 0.7628118903563929

total clf loss: 288.5247845891863 total adv loss: 234.09230513870716

evaluating...

Confusion matrix, without normalization

[[ 41 221 32]

[ 31 3697 93]

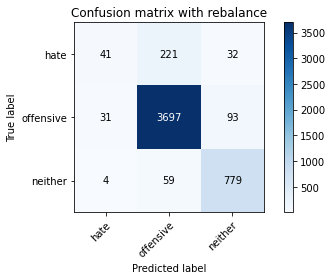
[ 4 59 779]]

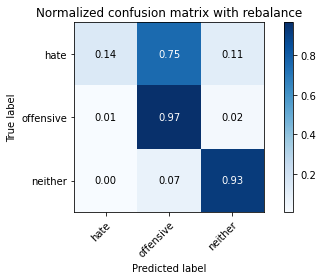
Normalized confusion matrix

[[0.13945578 0.75170068 0.10884354]

[0.00811306 0.96754776 0.02433918]

[0.00475059 0.07007126 0.92517815]]





precision recall f1-score support

hate 0.54 0.14 0.22 294

offensive 0.93 0.97 0.95 3821

neither 0.86 0.93 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.68 0.69 4957

weighted avg 0.89 0.91 0.90 4957

average eval\_loss: 0.24766655497613452, accuracy: 0.9112366350615292

training 2 epoch...

843.7937576770782 seconds used......

2 training finished! average train loss: 0.5832445827282725

total clf loss: 212.89350005611777 total adv loss: 182.84256272949278

evaluating...

Confusion matrix, without normalization

[[ 104 162 28]

[ 96 3638 87]

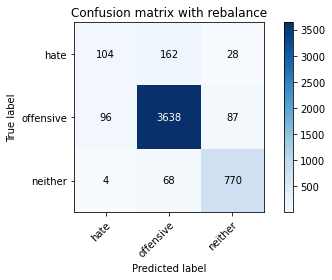
[ 4 68 770]]

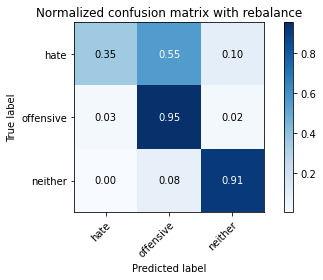
Normalized confusion matrix

[[0.3537415 0.55102041 0.0952381 ]

[0.02512431 0.95210678 0.02276891]

[0.00475059 0.0807601 0.91448931]]





precision recall f1-score support

hate 0.51 0.35 0.42 294

offensive 0.94 0.95 0.95 3821

neither 0.87 0.91 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.74 0.75 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.24370145105246094, accuracy: 0.9102279604599556

training 3 epoch...

843.7676479816437 seconds used......

3 training finished! average train loss: 0.4551384665913159

total clf loss: 164.52395019633695 total adv loss: 143.4867039443925

evaluating...

Confusion matrix, without normalization

[[ 138 134 22]

[ 126 3604 91]

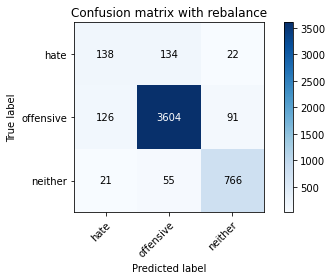
[ 21 55 766]]

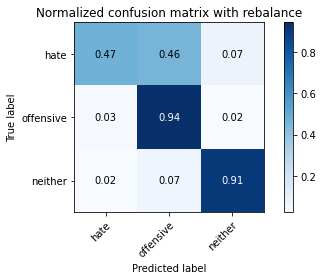
Normalized confusion matrix

[[0.46938776 0.45578231 0.07482993]

[0.03297566 0.94320858 0.02381576]

[0.02494062 0.06532067 0.90973872]]





precision recall f1-score support

hate 0.48 0.47 0.48 294

offensive 0.95 0.94 0.95 3821

neither 0.87 0.91 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.77 0.77 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.28209580724397976, accuracy: 0.9094210207786968

training 4 epoch...

843.7498672008514 seconds used......

4 training finished! average train loss: 0.32267846730856914

total clf loss: 116.72843250026926 total adv loss: 101.68430317391176

evaluating...

Confusion matrix, without normalization

[[ 141 141 12]

[ 128 3626 67]

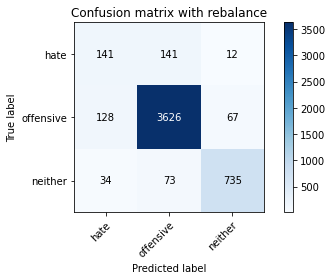
[ 34 73 735]]

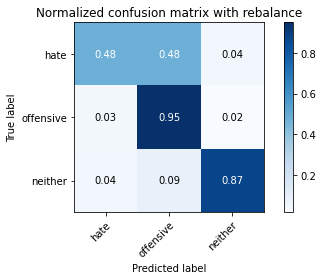
Normalized confusion matrix

[[0.47959184 0.47959184 0.04081633]

[0.03349908 0.94896624 0.01753468]

[0.04038005 0.08669834 0.87292162]]





precision recall f1-score support

hate 0.47 0.48 0.47 294

offensive 0.94 0.95 0.95 3821

neither 0.90 0.87 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.77 0.77 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.3070878889205145, accuracy: 0.9082106112568086

training 5 epoch...

843.7016549110413 seconds used......

5 training finished! average train loss: 0.20012884566048958

total clf loss: 71.4740131496801 total adv loss: 63.52690089395401

evaluating...

Confusion matrix, without normalization

[[ 121 149 24]

[ 131 3604 86]

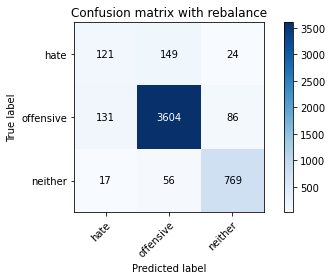
[ 17 56 769]]

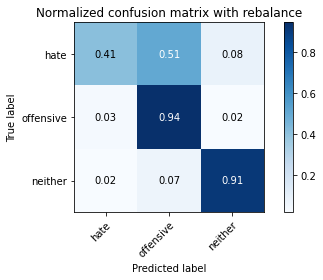
Normalized confusion matrix

[[0.41156463 0.50680272 0.08163265]

[0.03428422 0.94320858 0.0225072 ]

[0.02019002 0.06650831 0.91330166]]





precision recall f1-score support

hate 0.45 0.41 0.43 294

offensive 0.95 0.94 0.94 3821

neither 0.87 0.91 0.89 842

accuracy 0.91 4957

macro avg 0.76 0.76 0.76 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.40922663187971453, accuracy: 0.9065967318942909

4 run......

training 1 epoch...

844.0069546699524 seconds used......

1 training finished! average train loss: 0.7514583762629979

total clf loss: 284.34223464131355 total adv loss: 230.55223717354238

evaluating...

Confusion matrix, without normalization

[[ 77 191 26]

[ 72 3682 67]

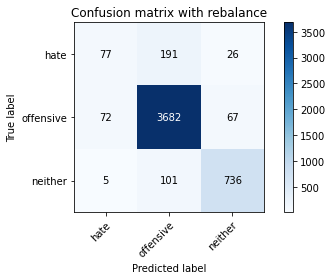
[ 5 101 736]]

Normalized confusion matrix

[[0.26190476 0.64965986 0.08843537]

[0.01884323 0.96362209 0.01753468]

[0.00593824 0.11995249 0.87410926]]





precision recall f1-score support

hate 0.50 0.26 0.34 294

offensive 0.93 0.96 0.94 3821

neither 0.89 0.87 0.88 842

accuracy 0.91 4957

macro avg 0.77 0.70 0.72 4957

weighted avg 0.89 0.91 0.90 4957

average eval\_loss: 0.25759719362512473, accuracy: 0.9067984668146056

training 2 epoch...

844.0090343952179 seconds used......

2 training finished! average train loss: 0.5856561454373502

total clf loss: 214.71171662956476 total adv loss: 183.1295899618417

evaluating...

Confusion matrix, without normalization

[[ 83 178 33]

[ 67 3656 98]

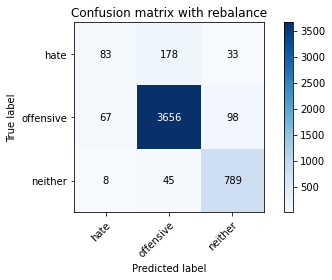
[ 8 45 789]]

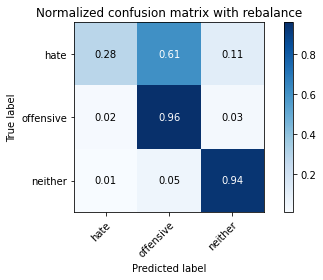
Normalized confusion matrix

[[0.28231293 0.60544218 0.1122449 ]

[0.01753468 0.95681759 0.02564774]

[0.00950119 0.05344418 0.93705463]]





precision recall f1-score support

hate 0.53 0.28 0.37 294

offensive 0.94 0.96 0.95 3821

neither 0.86 0.94 0.90 842

accuracy 0.91 4957

macro avg 0.78 0.73 0.74 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.24367593913414184, accuracy: 0.9134557191849909

training 3 epoch...

843.8604583740234 seconds used......

3 training finished! average train loss: 0.4588737791467969

total clf loss: 166.72945837676525 total adv loss: 144.2366654034704

evaluating...

Confusion matrix, without normalization

[[ 105 172 17]

[ 72 3682 67]

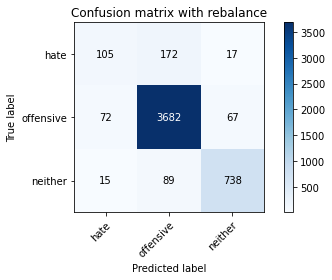
[ 15 89 738]]

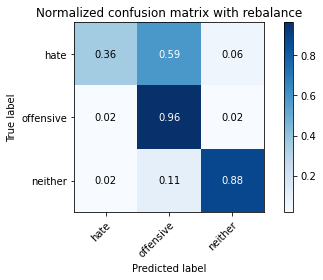
Normalized confusion matrix

[[0.35714286 0.58503401 0.05782313]

[0.01884323 0.96362209 0.01753468]

[0.01781473 0.10570071 0.87648456]]





precision recall f1-score support

hate 0.55 0.36 0.43 294

offensive 0.93 0.96 0.95 3821

neither 0.90 0.88 0.89 842

accuracy 0.91 4957

macro avg 0.79 0.73 0.76 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.25420013058840507, accuracy: 0.9128505144240469

training 4 epoch...

844.0087230205536 seconds used......

4 training finished! average train loss: 0.3254740620632806

total clf loss: 117.41282416169997 total adv loss: 102.72872251534136

evaluating...

Confusion matrix, without normalization

[[ 97 169 28]

[ 78 3651 92]

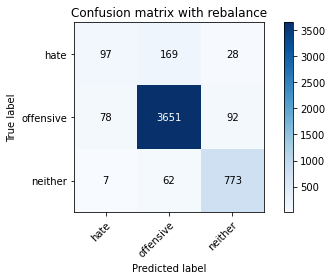
[ 7 62 773]]

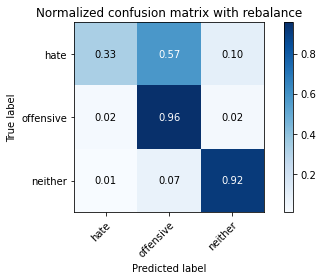
Normalized confusion matrix

[[0.32993197 0.57482993 0.0952381 ]

[0.0204135 0.95550903 0.02407747]

[0.00831354 0.0736342 0.91805226]]





precision recall f1-score support

hate 0.53 0.33 0.41 294

offensive 0.94 0.96 0.95 3821

neither 0.87 0.92 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.73 0.75 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.29757411026437164, accuracy: 0.912043574742788

training 5 epoch...

843.823591709137 seconds used......

5 training finished! average train loss: 0.18404564299553092

total clf loss: 67.96320538123837 total adv loss: 57.30503622408287

evaluating...

Confusion matrix, without normalization

[[ 115 162 17]

[ 151 3596 74]

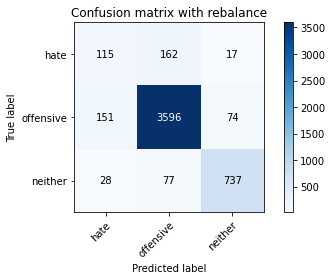
[ 28 77 737]]

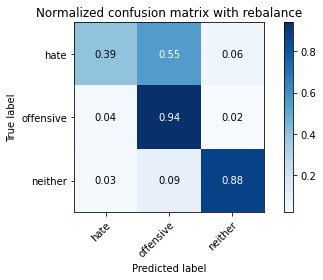
Normalized confusion matrix

[[0.39115646 0.55102041 0.05782313]

[0.03951845 0.94111489 0.01936666]

[0.03325416 0.09144893 0.87529691]]





precision recall f1-score support

hate 0.39 0.39 0.39 294

offensive 0.94 0.94 0.94 3821

neither 0.89 0.88 0.88 842

accuracy 0.90 4957

macro avg 0.74 0.74 0.74 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.3684237054793254, accuracy: 0.8973169255598143

5 run......

training 1 epoch...

843.8095235824585 seconds used......

1 training finished! average train loss: 0.7587042295208741

total clf loss: 287.1744337938726 total adv loss: 232.73008136451244

evaluating...

Confusion matrix, without normalization

[[ 18 231 45]

[ 14 3703 104]

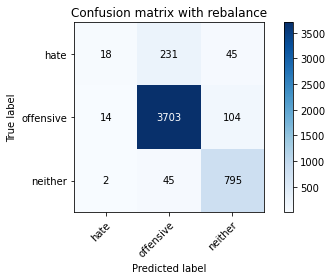
[ 2 45 795]]

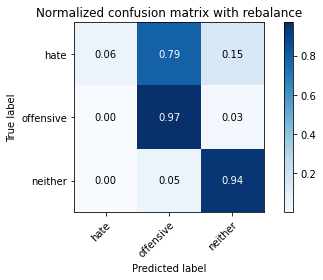
Normalized confusion matrix

[[0.06122449 0.78571429 0.15306122]

[0.00366396 0.96911803 0.02721801]

[0.0023753 0.05344418 0.94418052]]





precision recall f1-score support

hate 0.53 0.06 0.11 294

offensive 0.93 0.97 0.95 3821

neither 0.84 0.94 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.66 0.65 4957

weighted avg 0.89 0.91 0.89 4957

average eval\_loss: 0.2525673844821511, accuracy: 0.9110349001412145

training 2 epoch...

843.8616983890533 seconds used......

2 training finished! average train loss: 0.5889501760435861

total clf loss: 214.85765444859862 total adv loss: 184.69045994710177

evaluating...

Confusion matrix, without normalization

[[ 126 143 25]

[ 107 3621 93]

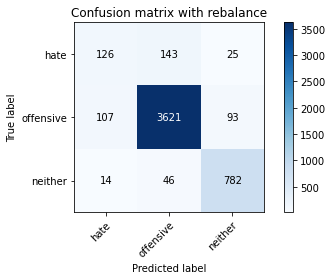
[ 14 46 782]]

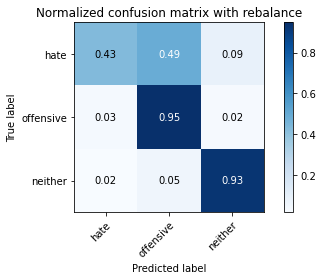
Normalized confusion matrix

[[0.42857143 0.48639456 0.08503401]

[0.02800314 0.94765768 0.02433918]

[0.01662708 0.05463183 0.92874109]]





precision recall f1-score support

hate 0.51 0.43 0.47 294

offensive 0.95 0.95 0.95 3821

neither 0.87 0.93 0.90 842

accuracy 0.91 4957

macro avg 0.78 0.77 0.77 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.2454058590585426, accuracy: 0.9136574541053056

training 3 epoch...

843.820508480072 seconds used......

3 training finished! average train loss: 0.45688754615313826

total clf loss: 165.65428453311324 total adv loss: 143.78908054996282

evaluating...

Confusion matrix, without normalization

[[ 42 233 19]

[ 22 3723 76]

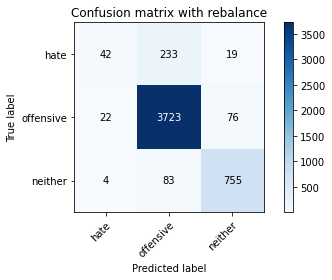
[ 4 83 755]]

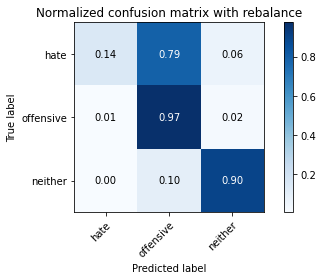
Normalized confusion matrix

[[0.14285714 0.79251701 0.06462585]

[0.00575766 0.97435226 0.01989008]

[0.00475059 0.09857482 0.89667458]]





precision recall f1-score support

hate 0.62 0.14 0.23 294

offensive 0.92 0.97 0.95 3821

neither 0.89 0.90 0.89 842

accuracy 0.91 4957

macro avg 0.81 0.67 0.69 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.2698974857471072, accuracy: 0.9118418398224732

training 4 epoch...

843.983512878418 seconds used......

4 training finished! average train loss: 0.323008789037328

total clf loss: 116.106469940627 total adv loss: 102.15912444551941

evaluating...

Confusion matrix, without normalization

[[ 80 189 25]

[ 63 3667 91]

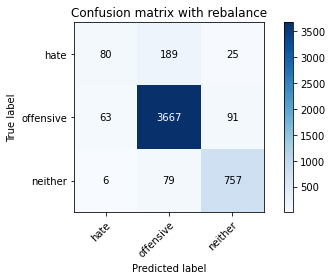
[ 6 79 757]]

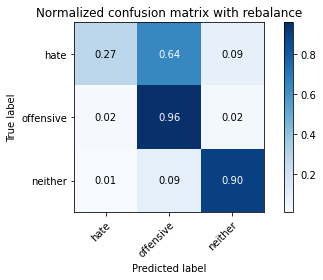
Normalized confusion matrix

[[0.27210884 0.64285714 0.08503401]

[0.01648783 0.95969641 0.02381576]

[0.00712589 0.09382423 0.89904988]]





precision recall f1-score support

hate 0.54 0.27 0.36 294

offensive 0.93 0.96 0.95 3821

neither 0.87 0.90 0.88 842

accuracy 0.91 4957

macro avg 0.78 0.71 0.73 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.2996469103541571, accuracy: 0.908614081097438

training 5 epoch...

844.4827990531921 seconds used......

5 training finished! average train loss: 0.20034171668756304

total clf loss: 70.99212099856231 total adv loss: 63.873431264277315

evaluating...

Confusion matrix, without normalization

[[ 113 163 18]

[ 113 3636 72]

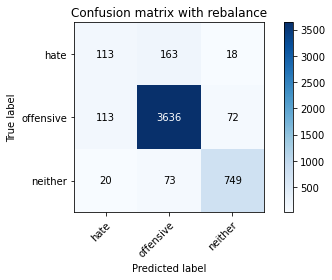
[ 20 73 749]]

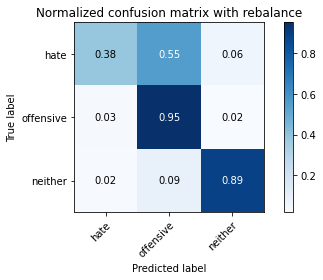
Normalized confusion matrix

[[0.38435374 0.55442177 0.06122449]

[0.02957341 0.95158336 0.01884323]

[0.02375297 0.08669834 0.88954869]]





precision recall f1-score support

hate 0.46 0.38 0.42 294

offensive 0.94 0.95 0.95 3821

neither 0.89 0.89 0.89 842

accuracy 0.91 4957

macro avg 0.76 0.74 0.75 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.34642464022768, accuracy: 0.9074036715755497

1 run......

Downloading: 100%

411/411 [00:00<00:00, 11.4kB/s]

Downloading: 100%

263M/263M [00:09<00:00, 18.4MB/s]

training 1 epoch...

847.3757886886597 seconds used......

1 training finished! average train loss: 0.7620315788719323

total clf loss: 290.0917943455279 total adv loss: 232.9217653274536

evaluating...

Confusion matrix, without normalization

[[ 14 244 36]

[ 7 3707 107]

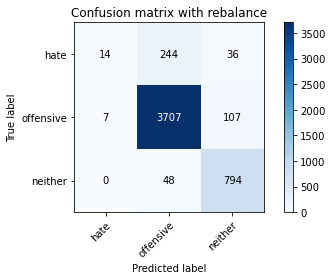
[ 0 48 794]]

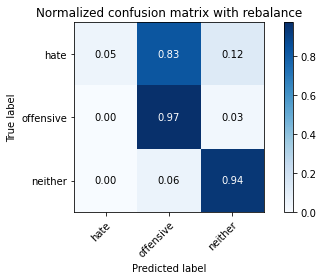
Normalized confusion matrix

[[0.04761905 0.82993197 0.12244898]

[0.00183198 0.97016488 0.02800314]

[0. 0.05700713 0.94299287]]





precision recall f1-score support

hate 0.67 0.05 0.09 294

offensive 0.93 0.97 0.95 3821

neither 0.85 0.94 0.89 842

accuracy 0.91 4957

macro avg 0.81 0.65 0.64 4957

weighted avg 0.90 0.91 0.89 4957

average eval\_loss: 0.2442061060048159, accuracy: 0.9108331652208997

training 2 epoch...

847.0220017433167 seconds used......

2 training finished! average train loss: 0.5783305712862902

total clf loss: 211.51468146406114 total adv loss: 181.09462248021737

evaluating...

Confusion matrix, without normalization

[[ 114 161 19]

[ 109 3629 83]

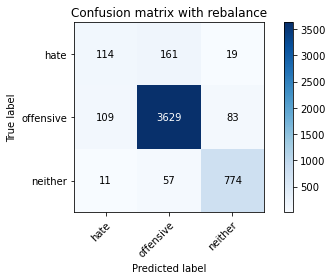
[ 11 57 774]]

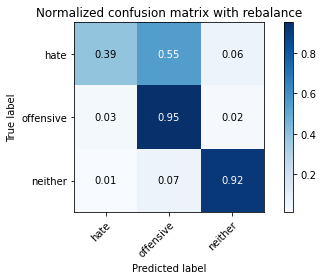
Normalized confusion matrix

[[0.3877551 0.54761905 0.06462585]

[0.02852656 0.94975137 0.02172206]

[0.01306413 0.06769596 0.9192399 ]]





precision recall f1-score support

hate 0.49 0.39 0.43 294

offensive 0.94 0.95 0.95 3821

neither 0.88 0.92 0.90 842

accuracy 0.91 4957

macro avg 0.77 0.75 0.76 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.24766567465473688, accuracy: 0.9112366350615292

training 3 epoch...

847.0512099266052 seconds used......

3 training finished! average train loss: 0.4473448292194535

total clf loss: 162.69999869773164 total adv loss: 140.53303620312363

evaluating...

Confusion matrix, without normalization

[[ 81 190 23]

[ 45 3707 69]

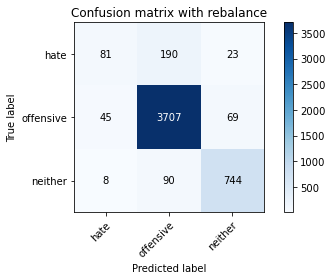
[ 8 90 744]]

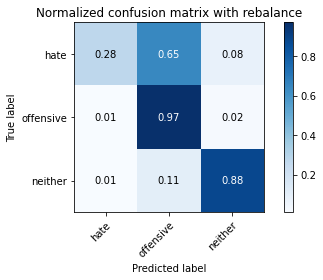
Normalized confusion matrix

[[0.2755102 0.6462585 0.07823129]

[0.01177702 0.97016488 0.0180581 ]

[0.00950119 0.10688836 0.88361045]]





precision recall f1-score support

hate 0.60 0.28 0.38 294

offensive 0.93 0.97 0.95 3821

neither 0.89 0.88 0.89 842

accuracy 0.91 4957

macro avg 0.81 0.71 0.74 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.2620982357006024, accuracy: 0.9142626588662498

training 4 epoch...

846.9482314586639 seconds used......

4 training finished! average train loss: 0.3096150715023067

total clf loss: 112.39818830613513 total adv loss: 97.36998128920095

evaluating...

Confusion matrix, without normalization

[[ 82 189 23]

[ 56 3685 80]

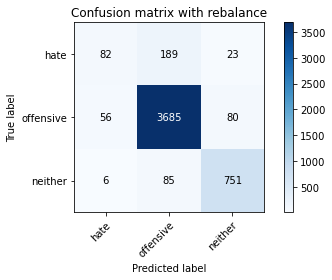
[ 6 85 751]]

Normalized confusion matrix

[[0.27891156 0.64285714 0.07823129]

[0.01465585 0.96440722 0.02093693]

[0.00712589 0.10095012 0.89192399]]





precision recall f1-score support

hate 0.57 0.28 0.37 294

offensive 0.93 0.96 0.95 3821

neither 0.88 0.89 0.89 842

accuracy 0.91 4957

macro avg 0.79 0.71 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.31142578749624533, accuracy: 0.9114383699818439

training 5 epoch...

847.4972369670868 seconds used......

5 training finished! average train loss: 0.1869188608074065

total clf loss: 68.00961657246808 total adv loss: 58.70694659657602

evaluating...

Confusion matrix, without normalization

[[ 67 199 28]

[ 45 3664 112]

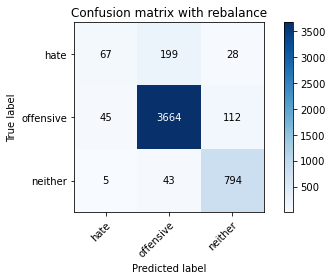
[ 5 43 794]]

Normalized confusion matrix

[[0.22789116 0.67687075 0.0952381 ]

[0.01177702 0.95891128 0.0293117 ]

[0.00593824 0.05106888 0.94299287]]





precision recall f1-score support

hate 0.57 0.23 0.33 294

offensive 0.94 0.96 0.95 3821

neither 0.85 0.94 0.89 842

accuracy 0.91 4957

macro avg 0.79 0.71 0.72 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.40214369298783204, accuracy: 0.9128505144240469

training 6 epoch...

846.7432346343994 seconds used......

6 training finished! average train loss: 0.10966858972445945

total clf loss: 41.133059663159656 total adv loss: 33.82909066240063

evaluating...

Confusion matrix, without normalization

[[ 81 194 19]

[ 76 3677 68]

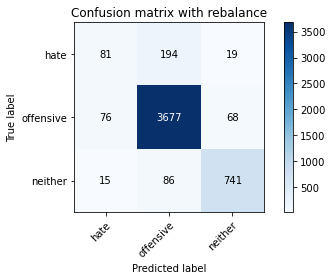
[ 15 86 741]]

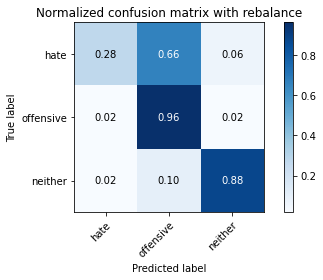
Normalized confusion matrix

[[0.2755102 0.65986395 0.06462585]

[0.01989008 0.96231353 0.01779639]

[0.01781473 0.10213777 0.88004751]]





precision recall f1-score support

hate 0.47 0.28 0.35 294

offensive 0.93 0.96 0.95 3821

neither 0.89 0.88 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.71 0.73 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.4350527565351982, accuracy: 0.9076054064958644

training 7 epoch...

847.05118060112 seconds used......

7 training finished! average train loss: 0.077535503718419

total clf loss: 29.23154702220927 total adv loss: 23.841836253193833

evaluating...

Confusion matrix, without normalization

[[ 95 182 17]

[ 84 3671 66]

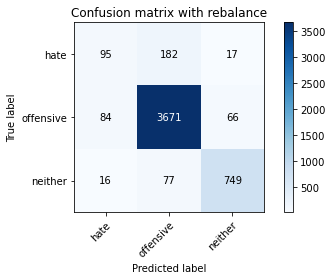
[ 16 77 749]]

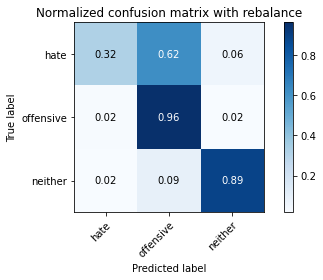
Normalized confusion matrix

[[0.32312925 0.61904762 0.05782313]

[0.02198377 0.96074326 0.01727297]

[0.01900238 0.09144893 0.88954869]]





precision recall f1-score support

hate 0.49 0.32 0.39 294

offensive 0.93 0.96 0.95 3821

neither 0.90 0.89 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.72 0.74 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.4980004384536392, accuracy: 0.9108331652208997

training 8 epoch...

848.0021133422852 seconds used......

8 training finished! average train loss: 0.06054439736981147

total clf loss: 22.034114184029022 total adv loss: 19.01296401188756

evaluating...

Confusion matrix, without normalization

[[ 96 178 20]

[ 93 3654 74]

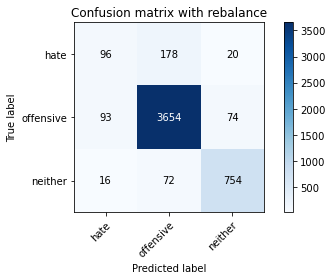
[ 16 72 754]]

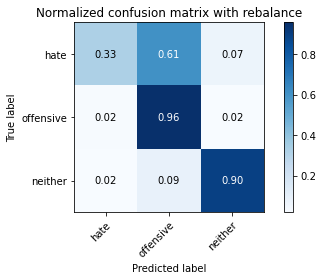
Normalized confusion matrix

[[0.32653061 0.60544218 0.06802721]

[0.02433918 0.95629416 0.01936666]

[0.01900238 0.08551069 0.89548694]]





precision recall f1-score support

hate 0.47 0.33 0.38 294

offensive 0.94 0.96 0.95 3821

neither 0.89 0.90 0.89 842

accuracy 0.91 4957

macro avg 0.76 0.73 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.4816109215153706, accuracy: 0.908614081097438

training 9 epoch...

847.4378814697266 seconds used......

9 training finished! average train loss: 0.05198863768004553

total clf loss: 19.207515432979562 total adv loss: 16.18260659010008

evaluating...

Confusion matrix, without normalization

[[ 137 138 19]

[ 189 3563 69]

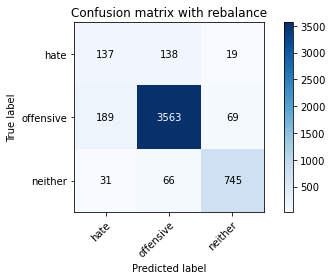
[ 31 66 745]]

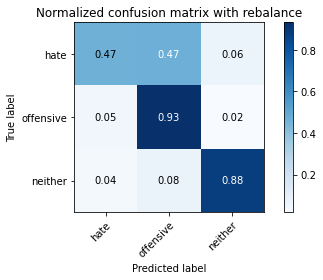
Normalized confusion matrix

[[0.46598639 0.46938776 0.06462585]

[0.04946349 0.93247841 0.0180581 ]

[0.0368171 0.0783848 0.8847981 ]]





precision recall f1-score support

hate 0.38 0.47 0.42 294

offensive 0.95 0.93 0.94 3821

neither 0.89 0.88 0.89 842

accuracy 0.90 4957

macro avg 0.74 0.76 0.75 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.5126396144569886, accuracy: 0.8967117207988703

training 10 epoch...

847.266693353653 seconds used......

10 training finished! average train loss: 0.0548302333096701

total clf loss: 19.00746741180774 total adv loss: 17.69206206999729

evaluating...

Confusion matrix, without normalization

[[ 99 170 25]

[ 73 3660 88]

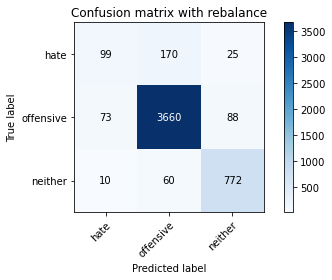
[ 10 60 772]]

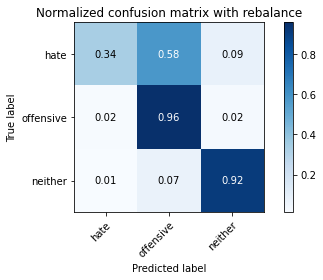
Normalized confusion matrix

[[0.33673469 0.57823129 0.08503401]

[0.01910495 0.95786443 0.02303062]

[0.01187648 0.07125891 0.91686461]]





precision recall f1-score support

hate 0.54 0.34 0.42 294

offensive 0.94 0.96 0.95 3821

neither 0.87 0.92 0.89 842

accuracy 0.91 4957

macro avg 0.79 0.74 0.75 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.5217818771535027, accuracy: 0.9140609239459351

2 run......

training 1 epoch...

846.9118156433105 seconds used......

1 training finished! average train loss: 0.7592950795792164

total clf loss: 289.00809010025114 total adv loss: 232.1063142362982

evaluating...

Confusion matrix, without normalization

[[ 51 215 28]

[ 46 3699 76]

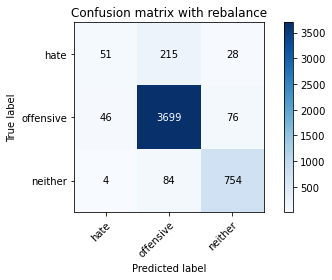
[ 4 84 754]]

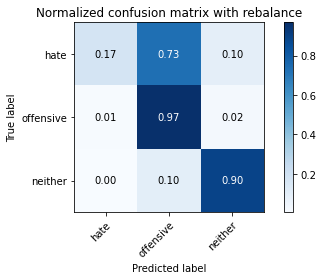
Normalized confusion matrix

[[0.17346939 0.73129252 0.0952381 ]

[0.01203873 0.96807119 0.01989008]

[0.00475059 0.09976247 0.89548694]]





precision recall f1-score support

hate 0.50 0.17 0.26 294

offensive 0.93 0.97 0.95 3821

neither 0.88 0.90 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.68 0.70 4957

weighted avg 0.89 0.91 0.90 4957

average eval\_loss: 0.2537966992916359, accuracy: 0.908614081097438

training 2 epoch...

847.1960790157318 seconds used......

2 training finished! average train loss: 0.5832824846879849

total clf loss: 213.1889332709834 total adv loss: 182.71364543912932

evaluating...

Confusion matrix, without normalization

[[ 29 238 27]

[ 14 3710 97]

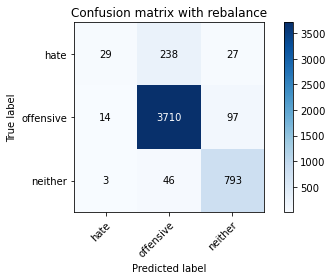
[ 3 46 793]]

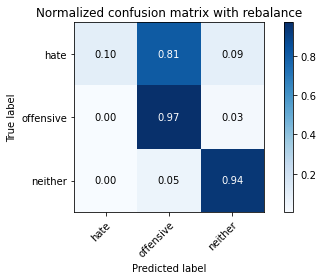
Normalized confusion matrix

[[0.09863946 0.80952381 0.09183673]

[0.00366396 0.97095001 0.02538602]

[0.00356295 0.05463183 0.94180523]]





precision recall f1-score support

hate 0.63 0.10 0.17 294

offensive 0.93 0.97 0.95 3821

neither 0.86 0.94 0.90 842

accuracy 0.91 4957

macro avg 0.81 0.67 0.67 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.235348888060769, accuracy: 0.9142626588662498

training 3 epoch...

846.9631814956665 seconds used......

3 training finished! average train loss: 0.44493472040818643

total clf loss: 161.3395457342267 total adv loss: 140.01784839329775

evaluating...

Confusion matrix, without normalization

[[ 97 174 23]

[ 72 3677 72]

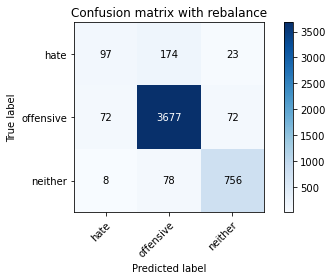
[ 8 78 756]]

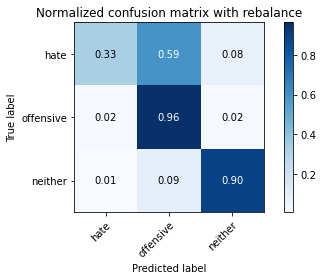
Normalized confusion matrix

[[0.32993197 0.59183673 0.07823129]

[0.01884323 0.96231353 0.01884323]

[0.00950119 0.09263658 0.89786223]]





precision recall f1-score support

hate 0.55 0.33 0.41 294

offensive 0.94 0.96 0.95 3821

neither 0.89 0.90 0.89 842

accuracy 0.91 4957

macro avg 0.79 0.73 0.75 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.2654163994311896, accuracy: 0.9138591890256204

training 4 epoch...

846.4522235393524 seconds used......

4 training finished! average train loss: 0.30839110460088237

total clf loss: 111.62080826517195 total adv loss: 97.15158377686748

evaluating...

Confusion matrix, without normalization

[[ 109 165 20]

[ 103 3649 69]

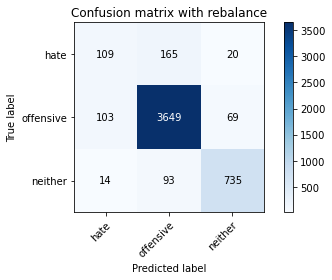
[ 14 93 735]]

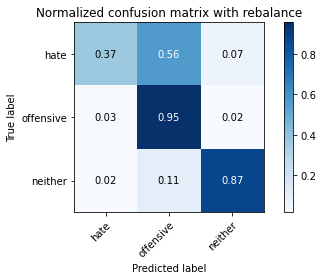
Normalized confusion matrix

[[0.3707483 0.56122449 0.06802721]

[0.02695629 0.95498561 0.0180581 ]

[0.01662708 0.11045131 0.87292162]]





precision recall f1-score support

hate 0.48 0.37 0.42 294

offensive 0.93 0.95 0.94 3821

neither 0.89 0.87 0.88 842

accuracy 0.91 4957

macro avg 0.77 0.73 0.75 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.31562327785321675, accuracy: 0.9063949969739762

training 5 epoch...