hate as off

alpha =0.05

1 run......

Downloading: 100%

411/411 [00:00<00:00, 10.6kB/s]

Downloading: 100%

263M/263M [00:11<00:00, 24.3MB/s]

training 1 epoch...

916.9651052951813 seconds used......

1 training finished! average train loss: 0.2807023734297423

total clf loss: 285.34679052978754 total adv loss: 137.8007342282217

evaluating...

Confusion matrix, without normalization

[[ 105 157 32]

[ 93 3644 84]

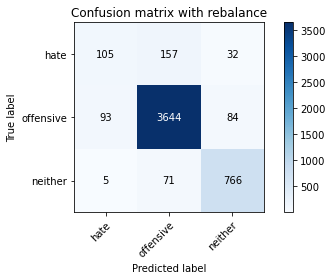
[ 5 71 766]]

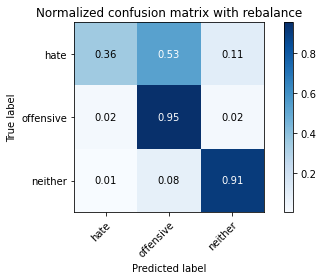
Normalized confusion matrix

[[0.35714286 0.53401361 0.10884354]

[0.02433918 0.95367705 0.02198377]

[0.00593824 0.08432304 0.90973872]]





precision recall f1-score support

hate 0.52 0.36 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 0.78 0.74 0.75 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.24176634771509037, accuracy: 0.9108331652208997

training 2 epoch...

917.2111003398895 seconds used......

2 training finished! average train loss: 0.20022804709057504

total clf loss: 202.6360983196646 total adv loss: 80.19751612030086

evaluating...

Confusion matrix, without normalization

[[ 93 179 22]

[ 60 3686 75]

[ 8 65 769]]

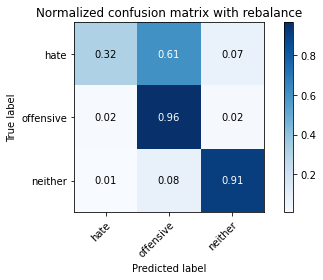
Normalized confusion matrix

[[0.31632653 0.60884354 0.07482993]

[0.0157027 0.96466893 0.01962837]

[0.00950119 0.07719715 0.91330166]]





precision recall f1-score support

hate 0.58 0.32 0.41 294

offensive 0.94 0.96 0.95 3821

neither 0.89 0.91 0.90 842

accuracy 0.92 4957

macro avg 0.80 0.73 0.75 4957

weighted avg 0.91 0.92 0.91 4957

average eval\_loss: 0.23958879125289498, accuracy: 0.917490417591285

training 3 epoch...

916.9365375041962 seconds used......

3 training finished! average train loss: 0.1429995131463417

total clf loss: 144.14637225633487 total adv loss: 45.817104393892805

evaluating...

Confusion matrix, without normalization

[[ 152 130 12]

[ 141 3612 68]

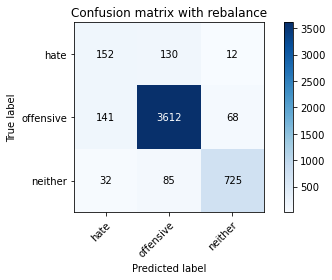
[ 32 85 725]]

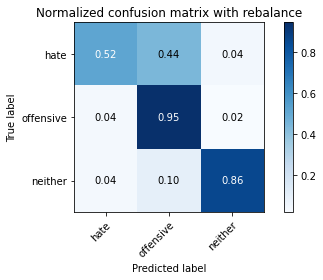
Normalized confusion matrix

[[0.5170068 0.44217687 0.04081633]

[0.03690133 0.94530228 0.01779639]

[0.03800475 0.10095012 0.86104513]]





precision recall f1-score support

hate 0.47 0.52 0.49 294

offensive 0.94 0.95 0.94 3821

neither 0.90 0.86 0.88 842

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: 0.29003439662084285, accuracy: 0.9055880572927174

training 4 epoch...

917.4239153862 seconds used......

4 training finished! average train loss: 0.0892025596807094

total clf loss: 89.7010326012969 total adv loss: 24.24186934010504

evaluating...

Confusion matrix, without normalization

[[ 60 214 20]

[ 41 3687 93]

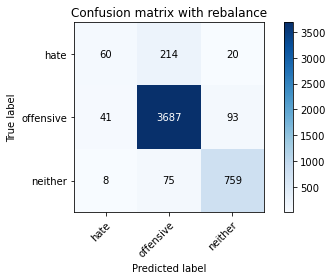
[ 8 75 759]]

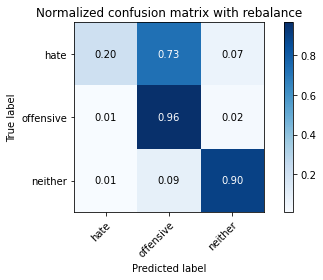
Normalized confusion matrix

[[0.20408163 0.72789116 0.06802721]

[0.01073018 0.96493065 0.02433918]

[0.00950119 0.08907363 0.90142518]]





precision recall f1-score support

hate 0.55 0.20 0.30 294

offensive 0.93 0.96 0.95 3821

neither 0.87 0.90 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.69 0.71 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.3621876895312227, accuracy: 0.9090175509380674

training 5 epoch...

916.5878496170044 seconds used......

5 training finished! average train loss: 0.04629898463890054

total clf loss: 46.64937539953098 total adv loss: 14.415653775569808

evaluating...

Confusion matrix, without normalization

[[ 116 160 18]

[ 136 3609 76]

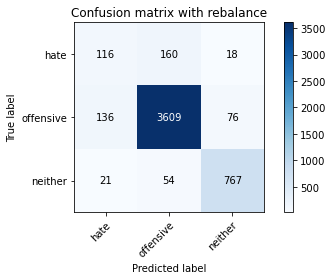
[ 21 54 767]]

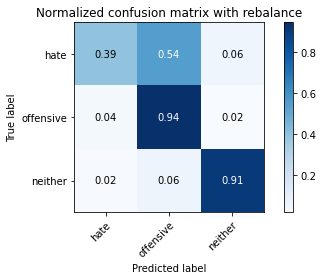
Normalized confusion matrix

[[0.39455782 0.54421769 0.06122449]

[0.03559278 0.94451714 0.01989008]

[0.02494062 0.06413302 0.91092637]]





precision recall f1-score support

hate 0.42 0.39 0.41 294

offensive 0.94 0.94 0.94 3821

neither 0.89 0.91 0.90 842

accuracy 0.91 4957

macro avg 0.75 0.75 0.75 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.4248795666387577, accuracy: 0.9061932620536615

2 run

1 run......

Downloading: 100%

411/411 [00:00<00:00, 14.2kB/s]

Downloading: 100%

263M/263M [00:23<00:00, 12.5MB/s]

training 1 epoch...

928.0684831142426 seconds used......

1 training finished! average train loss: 0.27268105717713853

total clf loss: 277.02103808149695 total adv loss: 130.4285873011686

evaluating...

Confusion matrix, without normalization

[[ 24 236 34]

[ 17 3699 105]

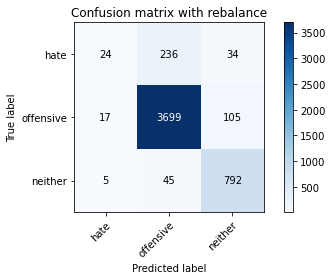
[ 5 45 792]]

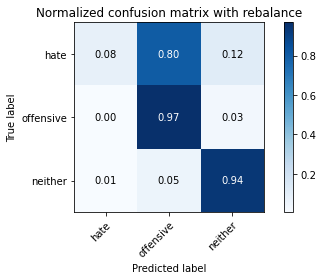
Normalized confusion matrix

[[0.08163265 0.80272109 0.11564626]

[0.0044491 0.96807119 0.02747972]

[0.00593824 0.05344418 0.94061758]]





precision recall f1-score support

hate 0.52 0.08 0.14 294

offensive 0.93 0.97 0.95 3821

neither 0.85 0.94 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.66 0.66 4957

weighted avg 0.89 0.91 0.89 4957

average eval\_loss: 0.24760676465267617, accuracy: 0.9108331652208997

training 2 epoch...

928.3840413093567 seconds used......

2 training finished! average train loss: 0.1959051365697486

total clf loss: 198.19505414180458 total adv loss: 77.1431711325422

evaluating...

Confusion matrix, without normalization

[[ 60 210 24]

[ 41 3690 90]

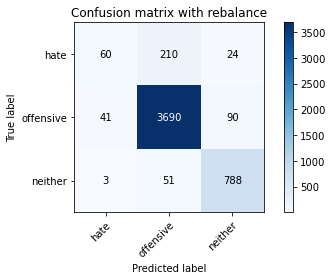
[ 3 51 788]]

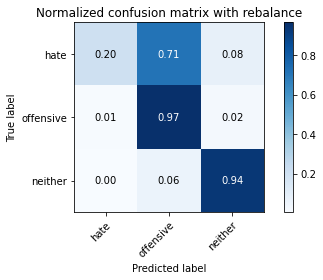
Normalized confusion matrix

[[0.20408163 0.71428571 0.08163265]

[0.01073018 0.96571578 0.02355404]

[0.00356295 0.06057007 0.93586698]]





precision recall f1-score support

hate 0.58 0.20 0.30 294

offensive 0.93 0.97 0.95 3821

neither 0.87 0.94 0.90 842

accuracy 0.92 4957

macro avg 0.79 0.70 0.72 4957

weighted avg 0.90 0.92 0.90 4957

average eval\_loss: 0.23825659340382704, accuracy: 0.915473068388138

training 3 epoch...

929.0049221515656 seconds used......

3 training finished! average train loss: 0.13644598512872752

total clf loss: 137.4705174509436 total adv loss: 42.32200960260525

evaluating...

Confusion matrix, without normalization

[[ 127 149 18]

[ 114 3635 72]

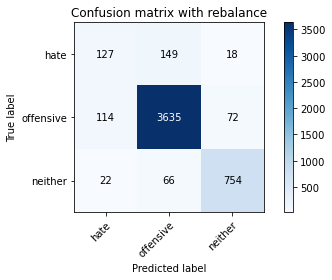
[ 22 66 754]]

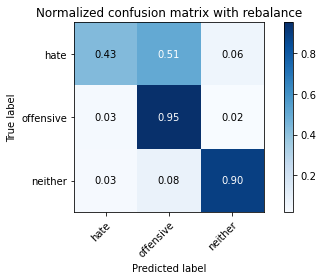
Normalized confusion matrix

[[0.43197279 0.50680272 0.06122449]

[0.02983512 0.95132164 0.01884323]

[0.02612827 0.0783848 0.89548694]]





precision recall f1-score support

hate 0.48 0.43 0.46 294

offensive 0.94 0.95 0.95 3821

neither 0.89 0.90 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.76 0.77 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.2668376412625695, accuracy: 0.9110349001412145

training 4 epoch...

928.2380051612854 seconds used......

4 training finished! average train loss: 0.08464921338124823

total clf loss: 85.22517929965397 total adv loss: 25.063191117696988

evaluating...

Confusion matrix, without normalization

[[ 127 141 26]

[ 172 3562 87]

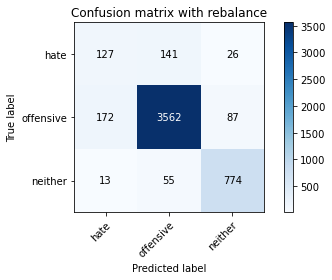
[ 13 55 774]]

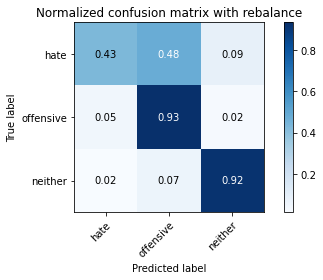
Normalized confusion matrix

[[0.43197279 0.47959184 0.08843537]

[0.04501439 0.9322167 0.02276891]

[0.01543943 0.06532067 0.9192399 ]]





precision recall f1-score support

hate 0.41 0.43 0.42 294

offensive 0.95 0.93 0.94 3821

neither 0.87 0.92 0.90 842

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: 0.35207891848192724, accuracy: 0.900342949364535

training 5 epoch...

927.5165829658508 seconds used......

5 training finished! average train loss: 0.039219165139324864

total clf loss: 39.4773127749213 total adv loss: 11.438019128483575

evaluating...

Confusion matrix, without normalization

[[ 116 164 14]

[ 165 3601 55]

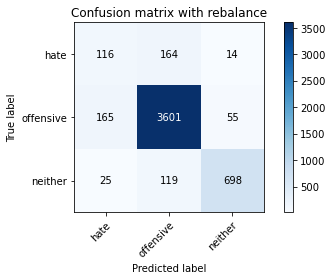
[ 25 119 698]]

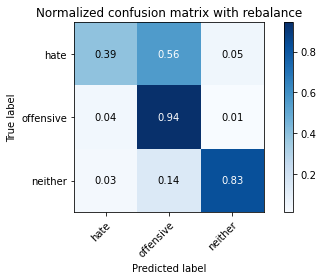
Normalized confusion matrix

[[0.39455782 0.55782313 0.04761905]

[0.04318241 0.94242345 0.01439414]

[0.02969121 0.14133017 0.82897862]]





precision recall f1-score support

hate 0.38 0.39 0.39 294

offensive 0.93 0.94 0.93 3821

neither 0.91 0.83 0.87 842

accuracy 0.89 4957

macro avg 0.74 0.72 0.73 4957

weighted avg 0.89 0.89 0.89 4957

average eval\_loss: 0.4625189005450774, accuracy: 0.8906596731894291

training 6 epoch...

927.3906617164612 seconds used......

6 training finished! average train loss: 0.026424385043949974

total clf loss: 26.751886917416414 total adv loss: 10.777939741117734

evaluating...

Confusion matrix, without normalization

[[ 126 154 14]

[ 146 3619 56]

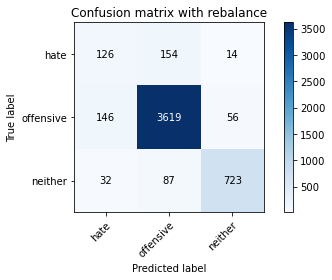
[ 32 87 723]]

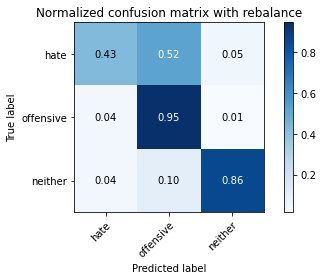
Normalized confusion matrix

[[0.42857143 0.52380952 0.04761905]

[0.03820989 0.94713426 0.01465585]

[0.03800475 0.10332542 0.85866983]]





precision recall f1-score support

hate 0.41 0.43 0.42 294

offensive 0.94 0.95 0.94 3821

neither 0.91 0.86 0.88 842

accuracy 0.90 4957

macro avg 0.75 0.74 0.75 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.46280096368032136, accuracy: 0.9013516239661086

training 7 epoch...

926.9220056533813 seconds used......

7 training finished! average train loss: 0.021565760789595486

total clf loss: 21.901879943539825 total adv loss: 10.172905801626257

evaluating...

Confusion matrix, without normalization

[[ 87 184 23]

[ 67 3657 97]

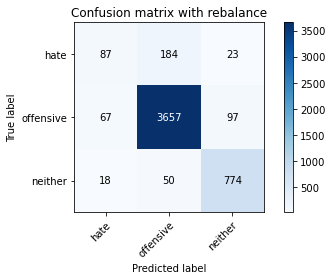
[ 18 50 774]]

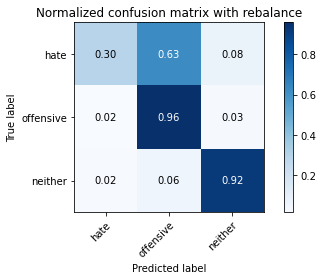
Normalized confusion matrix

[[0.29591837 0.62585034 0.07823129]

[0.01753468 0.9570793 0.02538602]

[0.02137767 0.05938242 0.9192399 ]]





precision recall f1-score support

hate 0.51 0.30 0.37 294

offensive 0.94 0.96 0.95 3821

neither 0.87 0.92 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.72 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.5325465858019133, accuracy: 0.9114383699818439

training 8 epoch...

926.7055311203003 seconds used......

8 training finished! average train loss: 0.016014209466414932

total clf loss: 16.257611364202603 total adv loss: 7.430312195133411

evaluating...

Confusion matrix, without normalization

[[ 97 184 13]

[ 101 3675 45]

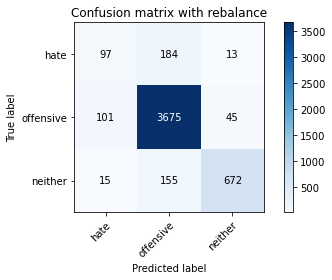
[ 15 155 672]]

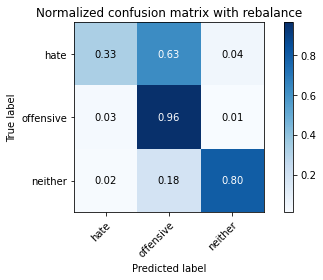
Normalized confusion matrix

[[0.32993197 0.62585034 0.04421769]

[0.02643287 0.96179011 0.01177702]

[0.01781473 0.18408551 0.79809976]]





precision recall f1-score support

hate 0.46 0.33 0.38 294

offensive 0.92 0.96 0.94 3821

neither 0.92 0.80 0.85 842

accuracy 0.90 4957

macro avg 0.76 0.70 0.73 4957

weighted avg 0.89 0.90 0.89 4957

average eval\_loss: 0.5511354690525702, accuracy: 0.8965099858785556

training 9 epoch...

926.9855618476868 seconds used......

9 training finished! average train loss: 0.016156447901395157

total clf loss: 16.344382788796793 total adv loss: 6.343728817121701

evaluating...

Confusion matrix, without normalization

[[ 145 133 16]

[ 262 3489 70]

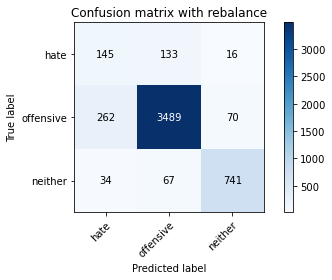
[ 34 67 741]]

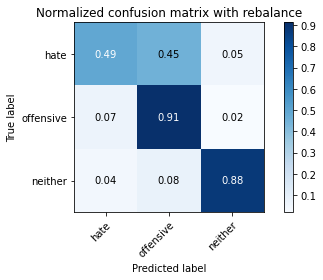
Normalized confusion matrix

[[0.49319728 0.45238095 0.05442177]

[0.06856844 0.91311175 0.01831981]

[0.04038005 0.07957245 0.88004751]]





precision recall f1-score support

hate 0.33 0.49 0.39 294

offensive 0.95 0.91 0.93 3821

neither 0.90 0.88 0.89 842

accuracy 0.88 4957

macro avg 0.72 0.76 0.74 4957

weighted avg 0.90 0.88 0.89 4957

average eval\_loss: 0.5866912553710497, accuracy: 0.8825902763768408

training 10 epoch...

926.7096571922302 seconds used......

10 training finished! average train loss: 0.014197054095857751

total clf loss: 14.432243899726018 total adv loss: 6.975325080523362

evaluating...

Confusion matrix, without normalization

[[ 92 179 23]

[ 81 3638 102]

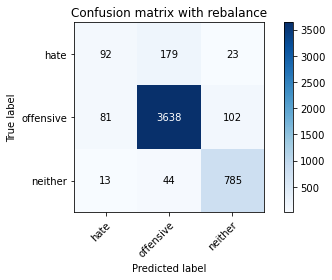
[ 13 44 785]]

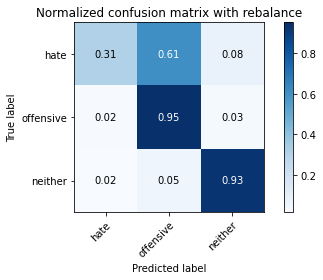
Normalized confusion matrix

[[0.31292517 0.60884354 0.07823129]

[0.02119864 0.95210678 0.02669458]

[0.01543943 0.05225653 0.93230404]]





precision recall f1-score support

hate 0.49 0.31 0.38 294

offensive 0.94 0.95 0.95 3821

neither 0.86 0.93 0.90 842

accuracy 0.91 4957

macro avg 0.77 0.73 0.74 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.6486547307331677, accuracy: 0.9108331652208997

2 run......

training 1 epoch...

927.2802286148071 seconds used......

1 training finished! average train loss: 0.2800252582209425

total clf loss: 284.4693462010473 total adv loss: 133.6858002138324

evaluating...

Confusion matrix, without normalization

[[ 32 222 40]

[ 24 3691 106]

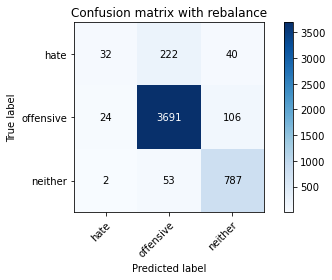
[ 2 53 787]]

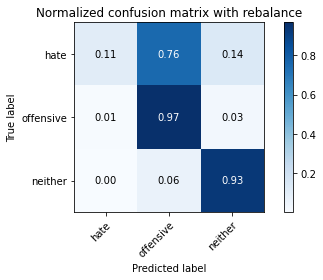
Normalized confusion matrix

[[0.10884354 0.75510204 0.13605442]

[0.00628108 0.96597749 0.02774143]

[0.0023753 0.06294537 0.93467933]]





precision recall f1-score support

hate 0.55 0.11 0.18 294

offensive 0.93 0.97 0.95 3821

neither 0.84 0.93 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.67 0.67 4957

weighted avg 0.89 0.91 0.89 4957

average eval\_loss: 0.25712481919195385, accuracy: 0.9098244906193262

training 2 epoch...

928.0609512329102 seconds used......

2 training finished! average train loss: 0.2002153808866898

total clf loss: 202.6427586292848 total adv loss: 80.58201533101965

evaluating...

Confusion matrix, without normalization

[[ 72 194 28]

[ 51 3665 105]

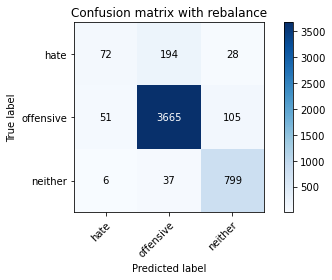
[ 6 37 799]]

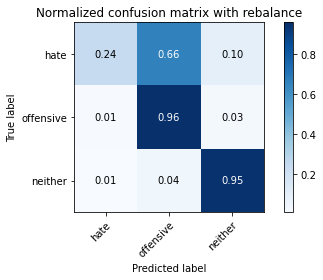
Normalized confusion matrix

[[0.24489796 0.65986395 0.0952381 ]

[0.01334729 0.95917299 0.02747972]

[0.00712589 0.04394299 0.94893112]]





precision recall f1-score support

hate 0.56 0.24 0.34 294

offensive 0.94 0.96 0.95 3821

neither 0.86 0.95 0.90 842

accuracy 0.92 4957

macro avg 0.79 0.72 0.73 4957

weighted avg 0.90 0.92 0.91 4957

average eval\_loss: 0.2516267318635308, accuracy: 0.9150695985475086

training 3 epoch...

928.6147086620331 seconds used......

3 training finished! average train loss: 0.14365702284539433

total clf loss: 144.83304529171437 total adv loss: 46.50557043039589

evaluating...

Confusion matrix, without normalization

[[ 104 173 17]

[ 84 3663 74]

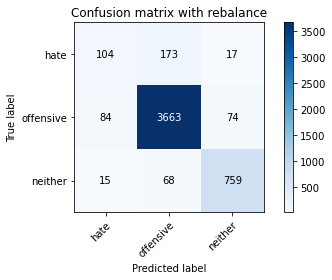
[ 15 68 759]]

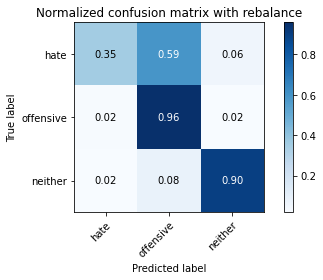
Normalized confusion matrix

[[0.3537415 0.58843537 0.05782313]

[0.02198377 0.95864957 0.01936666]

[0.01781473 0.0807601 0.90142518]]





precision recall f1-score support

hate 0.51 0.35 0.42 294

offensive 0.94 0.96 0.95 3821

neither 0.89 0.90 0.90 842

accuracy 0.91 4957

macro avg 0.78 0.74 0.75 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.27401294882198973, accuracy: 0.9130522493443615

training 4 epoch...

928.5657470226288 seconds used......

4 training finished! average train loss: 0.08787398727896113

total clf loss: 88.40110456710681 total adv loss: 24.602187396551017

evaluating...

Confusion matrix, without normalization

[[ 120 154 20]

[ 119 3633 69]

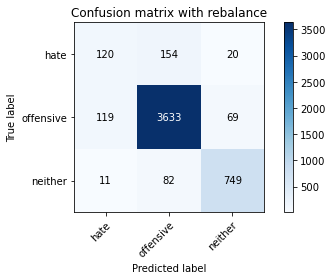
[ 11 82 749]]

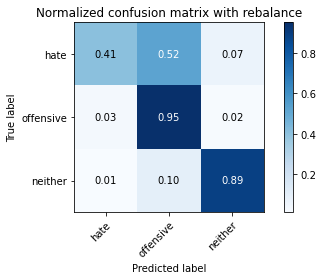
Normalized confusion matrix

[[0.40816327 0.52380952 0.06802721]

[0.03114368 0.95079822 0.0180581 ]

[0.01306413 0.09738717 0.88954869]]





precision recall f1-score support

hate 0.48 0.41 0.44 294

offensive 0.94 0.95 0.94 3821

neither 0.89 0.89 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.75 0.76 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.32589404962885554, accuracy: 0.9082106112568086

training 5 epoch...

928.2643985748291 seconds used......

5 training finished! average train loss: 0.049400197533208075

total clf loss: 49.804193903226405 total adv loss: 15.983959217499432

evaluating...

Confusion matrix, without normalization

[[ 101 174 19]

[ 83 3672 66]

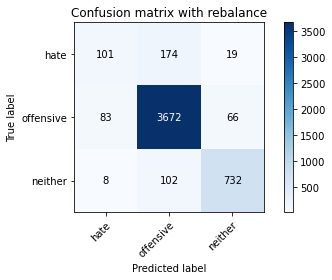
[ 8 102 732]]

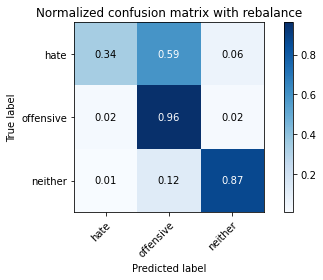
Normalized confusion matrix

[[0.34353741 0.59183673 0.06462585]

[0.02172206 0.96100497 0.01727297]

[0.00950119 0.12114014 0.86935867]]





precision recall f1-score support

hate 0.53 0.34 0.42 294

offensive 0.93 0.96 0.95 3821

neither 0.90 0.87 0.88 842

accuracy 0.91 4957

macro avg 0.78 0.72 0.75 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.3712228674761014, accuracy: 0.9088158160177526

training 6 epoch...

928.0129392147064 seconds used......

6 training finished! average train loss: 0.029353493198736796

total clf loss: 29.632485377340345 total adv loss: 10.276403076753013

evaluating...

Confusion matrix, without normalization

[[ 80 197 17]

[ 63 3698 60]

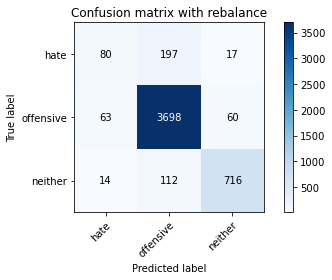
[ 14 112 716]]

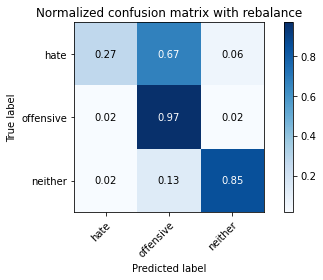
Normalized confusion matrix

[[0.27210884 0.67006803 0.05782313]

[0.01648783 0.96780947 0.0157027 ]

[0.01662708 0.13301663 0.85035629]]





precision recall f1-score support

hate 0.51 0.27 0.35 294

offensive 0.92 0.97 0.94 3821

neither 0.90 0.85 0.88 842

accuracy 0.91 4957

macro avg 0.78 0.70 0.73 4957

weighted avg 0.89 0.91 0.90 4957

average eval\_loss: 0.4489305414942552, accuracy: 0.9065967318942909

training 7 epoch...

927.7130124568939 seconds used......

7 training finished! average train loss: 0.017641038216476272

total clf loss: 17.769900596907974 total adv loss: 5.399812511868959

evaluating...

Confusion matrix, without normalization

[[ 112 164 18]

[ 121 3620 80]

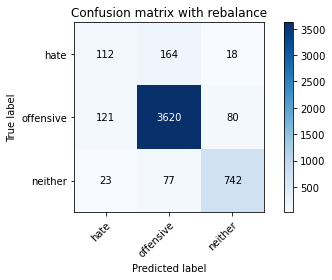
[ 23 77 742]]

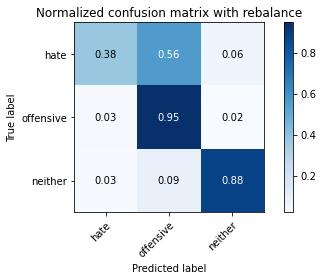
Normalized confusion matrix

[[0.38095238 0.55782313 0.06122449]

[0.0316671 0.94739597 0.02093693]

[0.02731591 0.09144893 0.88123515]]





precision recall f1-score support

hate 0.44 0.38 0.41 294

offensive 0.94 0.95 0.94 3821

neither 0.88 0.88 0.88 842

accuracy 0.90 4957

macro avg 0.75 0.74 0.74 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.4677162826160824, accuracy: 0.9025620334879968

training 8 epoch...

927.8132979869843 seconds used......

8 training finished! average train loss: 0.018451467667788282

total clf loss: 18.650315906379547 total adv loss: 6.929198799567757

evaluating...

Confusion matrix, without normalization

[[ 81 187 26]

[ 64 3666 91]

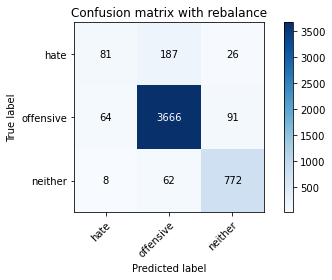
[ 8 62 772]]

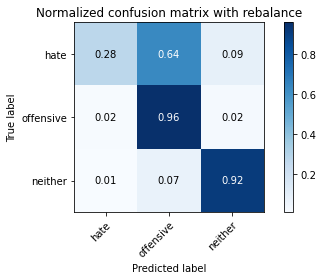
Normalized confusion matrix

[[0.2755102 0.63605442 0.08843537]

[0.01674954 0.9594347 0.02381576]

[0.00950119 0.0736342 0.91686461]]





precision recall f1-score support

hate 0.53 0.28 0.36 294

offensive 0.94 0.96 0.95 3821

neither 0.87 0.92 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.72 0.73 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.5359591037582329, accuracy: 0.9116401049021585

training 9 epoch...

927.9786310195923 seconds used......

9 training finished! average train loss: 0.013345030664613806

total clf loss: 13.601821900116192 total adv loss: 7.2710305036380305

evaluating...

Confusion matrix, without normalization

[[ 104 173 17]

[ 98 3662 61]

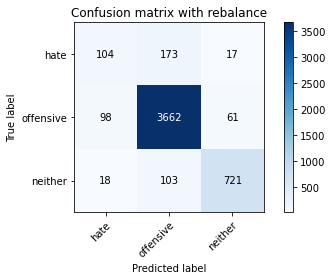
[ 18 103 721]]

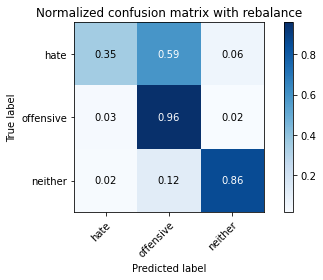
Normalized confusion matrix

[[0.3537415 0.58843537 0.05782313]

[0.02564774 0.95838786 0.01596441]

[0.02137767 0.12232779 0.85629454]]





precision recall f1-score support

hate 0.47 0.35 0.40 294

offensive 0.93 0.96 0.94 3821

neither 0.90 0.86 0.88 842

accuracy 0.91 4957

macro avg 0.77 0.72 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.5476881570681115, accuracy: 0.905184587452088

training 10 epoch...

927.9162836074829 seconds used......

10 training finished! average train loss: 0.012669393772002437

total clf loss: 12.809644481356372 total adv loss: 4.832117748539531

evaluating...

Confusion matrix, without normalization

[[ 126 151 17]

[ 174 3575 72]

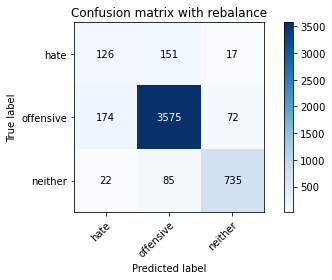
[ 22 85 735]]

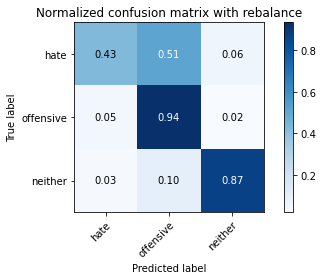
Normalized confusion matrix

[[0.42857143 0.51360544 0.05782313]

[0.04553782 0.93561895 0.01884323]

[0.02612827 0.10095012 0.87292162]]





precision recall f1-score support

hate 0.39 0.43 0.41 294

offensive 0.94 0.94 0.94 3821

neither 0.89 0.87 0.88 842

accuracy 0.89 4957

macro avg 0.74 0.75 0.74 4957

weighted avg 0.90 0.89 0.90 4957

average eval\_loss: 0.540905136457314, accuracy: 0.8948961065160379

3 run......

training 1 epoch...

926.6451127529144 seconds used......

1 training finished! average train loss: 0.27609820757061243

total clf loss: 280.5207153726369 total adv loss: 132.62587003223598

evaluating...

Confusion matrix, without normalization

[[ 92 172 30]

[ 78 3651 92]

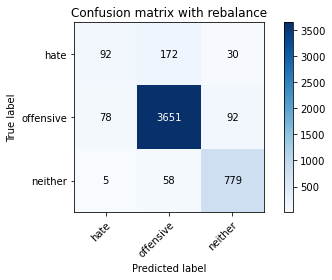
[ 5 58 779]]

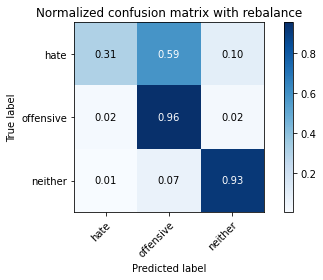
Normalized confusion matrix

[[0.31292517 0.58503401 0.10204082]

[0.0204135 0.95550903 0.02407747]

[0.00593824 0.06888361 0.92517815]]





precision recall f1-score support

hate 0.53 0.31 0.39 294

offensive 0.94 0.96 0.95 3821

neither 0.86 0.93 0.89 842

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: 0.24088911849614833, accuracy: 0.9122453096631027

training 2 epoch...

926.3017952442169 seconds used......

2 training finished! average train loss: 0.20114513931818487

total clf loss: 203.69451159611344 total adv loss: 83.17067279852927

evaluating...

Confusion matrix, without normalization

[[ 125 154 15]

[ 99 3657 65]

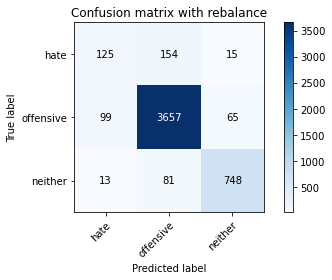
[ 13 81 748]]

Normalized confusion matrix

[[0.42517007 0.52380952 0.05102041]

[0.02590945 0.9570793 0.01701125]

[0.01543943 0.09619952 0.88836105]]





precision recall f1-score support

hate 0.53 0.43 0.47 294

offensive 0.94 0.96 0.95 3821

neither 0.90 0.89 0.90 842

accuracy 0.91 4957

macro avg 0.79 0.76 0.77 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.2483606898045588, accuracy: 0.9138591890256204

training 3 epoch...

925.9350955486298 seconds used......

3 training finished! average train loss: 0.14633688152647548

total clf loss: 147.64654016448185 total adv loss: 49.607075110398

evaluating...

Confusion matrix, without normalization

[[ 106 167 21]

[ 93 3648 80]

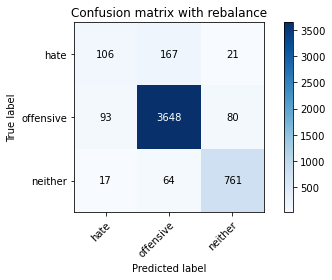
[ 17 64 761]]

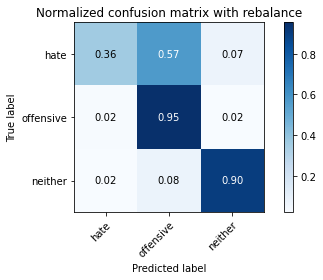
Normalized confusion matrix

[[0.36054422 0.56802721 0.07142857]

[0.02433918 0.95472389 0.02093693]

[0.02019002 0.0760095 0.90380048]]





precision recall f1-score support

hate 0.49 0.36 0.42 294

offensive 0.94 0.95 0.95 3821

neither 0.88 0.90 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: 0.26303500260016127, accuracy: 0.9108331652208997

training 4 epoch...

925.5024721622467 seconds used......

4 training finished! average train loss: 0.09179243200360948

total clf loss: 92.3416215943871 total adv loss: 25.67057647331967

evaluating...

Confusion matrix, without normalization

[[ 98 177 19]

[ 81 3673 67]

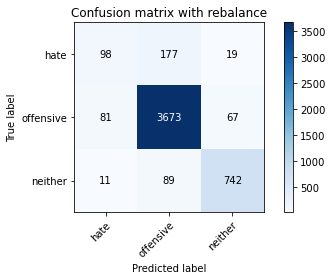
[ 11 89 742]]

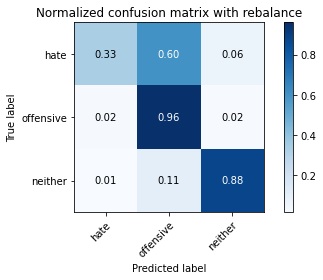
Normalized confusion matrix

[[0.33333333 0.60204082 0.06462585]

[0.02119864 0.96126668 0.01753468]

[0.01306413 0.10570071 0.88123515]]





precision recall f1-score support

hate 0.52 0.33 0.40 294

offensive 0.93 0.96 0.95 3821

neither 0.90 0.88 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.73 0.75 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.3313539290041175, accuracy: 0.9104296953802703

training 5 epoch...

925.997287273407 seconds used......

5 training finished! average train loss: 0.047815655431980804

total clf loss: 48.245665439317236 total adv loss: 16.25070485230026

evaluating...

Confusion matrix, without normalization

[[ 111 165 18]

[ 111 3627 83]

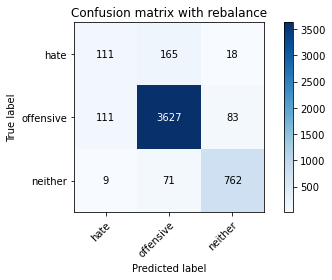
[ 9 71 762]]

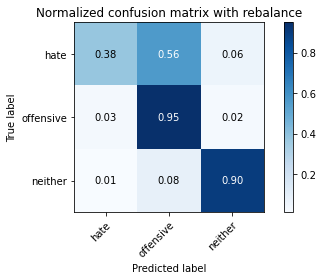
Normalized confusion matrix

[[0.37755102 0.56122449 0.06122449]

[0.02904999 0.94922795 0.02172206]

[0.01068884 0.08432304 0.90498812]]





precision recall f1-score support

hate 0.48 0.38 0.42 294

offensive 0.94 0.95 0.94 3821

neither 0.88 0.90 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.74 0.75 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.37352081615949095, accuracy: 0.9078071414161791

training 6 epoch...

926.1877191066742 seconds used......

6 training finished! average train loss: 0.030678804630937404

total clf loss: 31.09015431323496 total adv loss: 13.135602119844407

evaluating...

Confusion matrix, without normalization

[[ 88 189 17]

[ 66 3695 60]

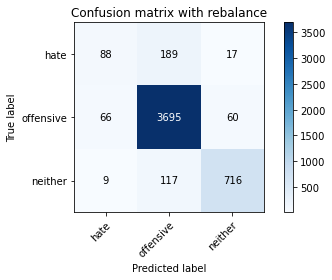
[ 9 117 716]]

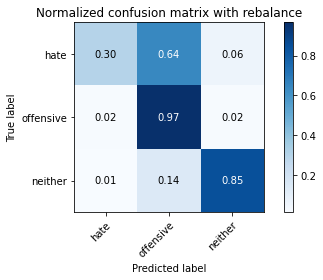
Normalized confusion matrix

[[0.29931973 0.64285714 0.05782313]

[0.01727297 0.96702434 0.0157027 ]

[0.01068884 0.13895487 0.85035629]]





precision recall f1-score support

hate 0.54 0.30 0.39 294

offensive 0.92 0.97 0.94 3821

neither 0.90 0.85 0.88 842

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: 0.49972910287021566, accuracy: 0.9076054064958644

training 7 epoch...

926.3966007232666 seconds used......

7 training finished! average train loss: 0.02183768136713356

total clf loss: 22.109389198565623 total adv loss: 8.92818598150825

evaluating...

Confusion matrix, without normalization

[[ 150 126 18]

[ 251 3499 71]

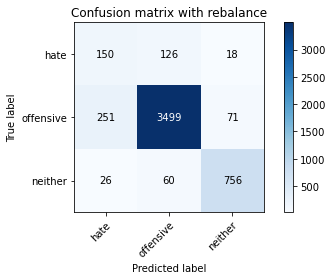
[ 26 60 756]]

Normalized confusion matrix

[[0.51020408 0.42857143 0.06122449]

[0.06568961 0.91572887 0.01858152]

[0.03087886 0.07125891 0.89786223]]





precision recall f1-score support

hate 0.35 0.51 0.42 294

offensive 0.95 0.92 0.93 3821

neither 0.89 0.90 0.90 842

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: 0.5706331991452551, accuracy: 0.8886423239862821

training 8 epoch...

926.5101203918457 seconds used......

8 training finished! average train loss: 0.018555677104933284

total clf loss: 18.794352870876537 total adv loss: 7.742424297293837

evaluating...

Confusion matrix, without normalization

[[ 134 147 13]

[ 129 3642 50]

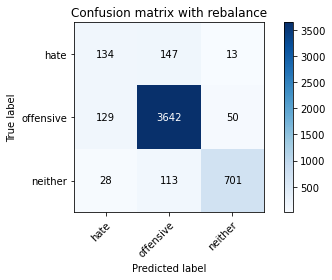
[ 28 113 701]]

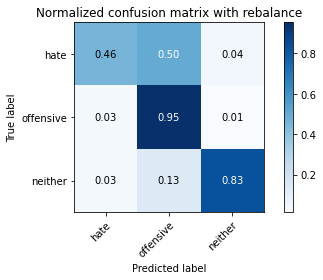
Normalized confusion matrix

[[0.45578231 0.5 0.04421769]

[0.0337608 0.95315362 0.01308558]

[0.03325416 0.13420428 0.83254157]]





precision recall f1-score support

hate 0.46 0.46 0.46 294

offensive 0.93 0.95 0.94 3821

neither 0.92 0.83 0.87 842

accuracy 0.90 4957

macro avg 0.77 0.75 0.76 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.5597470426539604, accuracy: 0.9031672382489409

training 9 epoch...

926.8796391487122 seconds used......

9 training finished! average train loss: 0.012991358529208605

total clf loss: 13.306074484495184 total adv loss: 8.372935994700583

evaluating...

Confusion matrix, without normalization

[[ 117 163 14]

[ 139 3628 54]

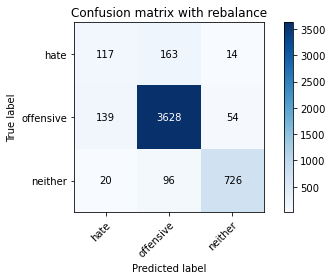
[ 20 96 726]]

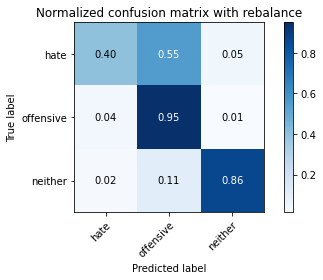
Normalized confusion matrix

[[0.39795918 0.55442177 0.04761905]

[0.03637791 0.94948966 0.01413243]

[0.02375297 0.11401425 0.86223278]]





precision recall f1-score support

hate 0.42 0.40 0.41 294

offensive 0.93 0.95 0.94 3821

neither 0.91 0.86 0.89 842

accuracy 0.90 4957

macro avg 0.76 0.74 0.75 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.5733276880720305, accuracy: 0.9019568287270526

training 10 epoch...

926.1434881687164 seconds used......

10 training finished! average train loss: 0.011262636445845572

total clf loss: 11.410300033276144 total adv loss: 4.755292571914879

evaluating...

Confusion matrix, without normalization

[[ 82 192 20]

[ 62 3677 82]

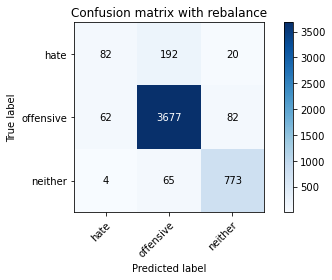
[ 4 65 773]]

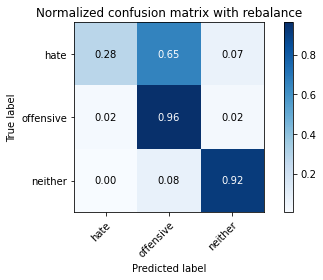
Normalized confusion matrix

[[0.27891156 0.65306122 0.06802721]

[0.01622612 0.96231353 0.02146035]

[0.00475059 0.07719715 0.91805226]]





precision recall f1-score support

hate 0.55 0.28 0.37 294

offensive 0.93 0.96 0.95 3821

neither 0.88 0.92 0.90 842

accuracy 0.91 4957

macro avg 0.79 0.72 0.74 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.5668117270459099, accuracy: 0.9142626588662498

4 run......

training 1 epoch...

927.5705978870392 seconds used......

1 training finished! average train loss: 0.27875170884879247

total clf loss: 283.1852535549551 total adv loss: 133.27116288500838

evaluating...

Confusion matrix, without normalization

[[ 112 162 20]

[ 101 3651 69]

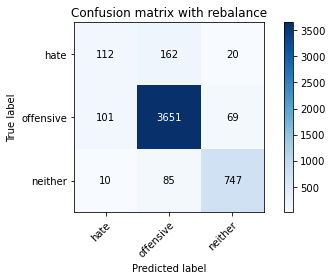
[ 10 85 747]]

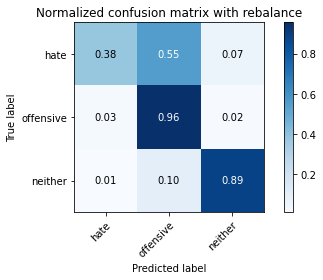
Normalized confusion matrix

[[0.38095238 0.55102041 0.06802721]

[0.02643287 0.95550903 0.0180581 ]

[0.01187648 0.10095012 0.8871734 ]]





precision recall f1-score support

hate 0.50 0.38 0.43 294

offensive 0.94 0.96 0.95 3821

neither 0.89 0.89 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.74 0.76 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.2544179175972878, accuracy: 0.9098244906193262

training 2 epoch...

5 run......

Downloading: 100%

411/411 [00:00<00:00, 6.32kB/s]

Downloading: 100%

263M/263M [00:12<00:00, 22.7MB/s]

training 1 epoch...

918.2940731048584 seconds used......

1 training finished! average train loss: 0.27537034246544806

total clf loss: 279.7992261592299 total adv loss: 132.63692672608886

evaluating...

Confusion matrix, without normalization

[[ 47 224 23]

[ 45 3698 78]

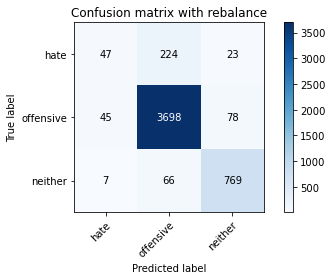
[ 7 66 769]]

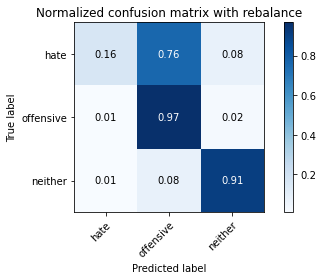
Normalized confusion matrix

[[0.15986395 0.76190476 0.07823129]

[0.01177702 0.96780947 0.0204135 ]

[0.00831354 0.0783848 0.91330166]]





precision recall f1-score support

hate 0.47 0.16 0.24 294

offensive 0.93 0.97 0.95 3821

neither 0.88 0.91 0.90 842

accuracy 0.91 4957

macro avg 0.76 0.68 0.69 4957

weighted avg 0.89 0.91 0.90 4957

average eval\_loss: 0.24396965159253486, accuracy: 0.910631430300585

training 2 epoch...

917.5125093460083 seconds used......

2 training finished! average train loss: 0.20083971115790547

total clf loss: 203.26262184977531 total adv loss: 80.5925705691916

evaluating...

Confusion matrix, without normalization

[[ 88 191 15]

[ 64 3684 73]

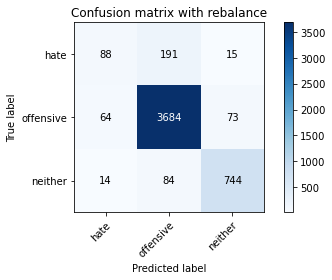
[ 14 84 744]]

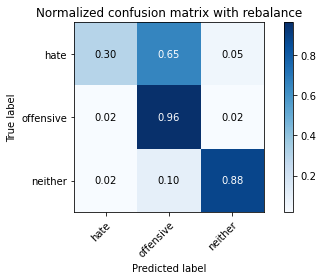
Normalized confusion matrix

[[0.29931973 0.64965986 0.05102041]

[0.01674954 0.96414551 0.01910495]

[0.01662708 0.09976247 0.88361045]]





precision recall f1-score support

hate 0.53 0.30 0.38 294

offensive 0.93 0.96 0.95 3821

neither 0.89 0.88 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.72 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.24020621352528612, accuracy: 0.9110349001412145

training 3 epoch...

917.6172573566437 seconds used......

3 training finished! average train loss: 0.14482896757685776

total clf loss: 146.04107815213501 total adv loss: 47.41484838842007

evaluating...

Confusion matrix, without normalization

[[ 115 164 15]

[ 91 3664 66]

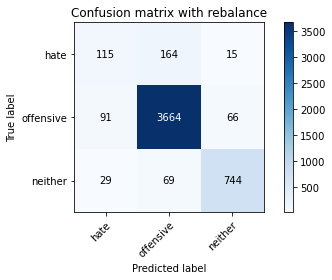
[ 29 69 744]]

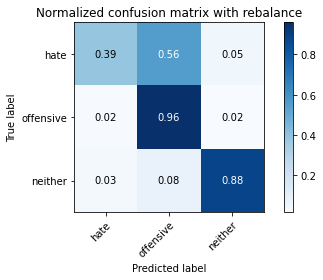
Normalized confusion matrix

[[0.39115646 0.55782313 0.05102041]

[0.02381576 0.95891128 0.01727297]

[0.03444181 0.08194774 0.88361045]]





precision recall f1-score support

hate 0.49 0.39 0.43 294

offensive 0.94 0.96 0.95 3821

neither 0.90 0.88 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.74 0.76 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.2751594357297666, accuracy: 0.9124470445834174

training 4 epoch...

917.6920444965363 seconds used......

4 training finished! average train loss: 0.0888064189140653

total clf loss: 89.3621945225168 total adv loss: 25.324535667507007

evaluating...

Confusion matrix, without normalization

[[ 91 187 16]

[ 72 3690 59]

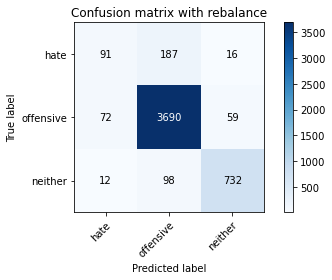
[ 12 98 732]]

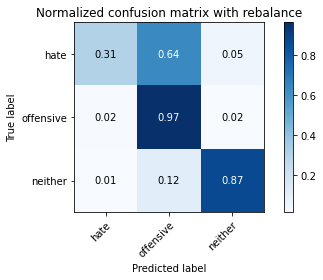
Normalized confusion matrix

[[0.30952381 0.63605442 0.05442177]

[0.01884323 0.96571578 0.01544098]

[0.01425178 0.11638955 0.86935867]]





precision recall f1-score support

hate 0.52 0.31 0.39 294

offensive 0.93 0.97 0.95 3821

neither 0.91 0.87 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: 0.31601722425999, accuracy: 0.9104296953802703

training 5 epoch...

917.5581562519073 seconds used......

5 training finished! average train loss: 0.045946374411165204

total clf loss: 46.29909178247908 total adv loss: 14.40576840050744

evaluating...

Confusion matrix, without normalization

[[ 106 164 24]

[ 90 3656 75]

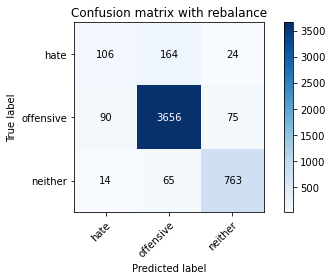
[ 14 65 763]]

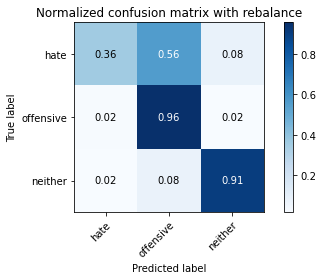
Normalized confusion matrix

[[0.36054422 0.55782313 0.08163265]

[0.02355404 0.95681759 0.01962837]

[0.01662708 0.07719715 0.90617577]]





precision recall f1-score support

hate 0.50 0.36 0.42 294

offensive 0.94 0.96 0.95 3821

neither 0.89 0.91 0.90 842

accuracy 0.91 4957

macro avg 0.78 0.74 0.76 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.38918941669039814, accuracy: 0.9128505144240469

training 6 epoch...

917.4703800678253 seconds used......

6 training finished! average train loss: 0.02826403012986409

total clf loss: 28.56733705739316 total adv loss: 10.588384439636684

evaluating...

Confusion matrix, without normalization

[[ 59 214 21]

[ 47 3702 72]

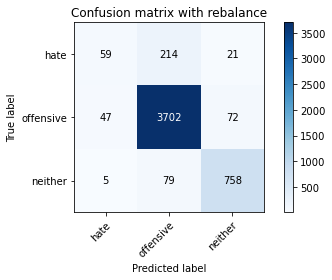
[ 5 79 758]]

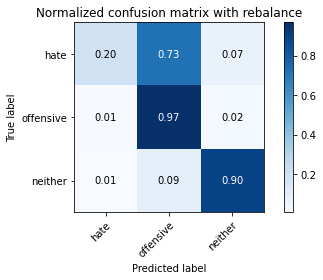
Normalized confusion matrix

[[0.20068027 0.72789116 0.07142857]

[0.01230044 0.96885632 0.01884323]

[0.00593824 0.09382423 0.90023753]]





precision recall f1-score support

hate 0.53 0.20 0.29 294

offensive 0.93 0.97 0.95 3821

neither 0.89 0.90 0.90 842

accuracy 0.91 4957

macro avg 0.78 0.69 0.71 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.46057269032298204, accuracy: 0.9116401049021585

training 7 epoch...

916.9503338336945 seconds used......

7 training finished! average train loss: 0.020737179194794966

total clf loss: 21.01671708550566 total adv loss: 8.908706446863562

evaluating...

Confusion matrix, without normalization

[[ 86 191 17]

[ 72 3688 61]

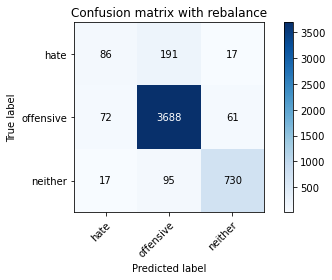
[ 17 95 730]]

Normalized confusion matrix

[[0.29251701 0.64965986 0.05782313]

[0.01884323 0.96519236 0.01596441]

[0.02019002 0.1128266 0.86698337]]





precision recall f1-score support

hate 0.49 0.29 0.37 294

offensive 0.93 0.97 0.95 3821

neither 0.90 0.87 0.88 842

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: 0.4982541707592733, accuracy: 0.908614081097438

training 8 epoch...

917.1949307918549 seconds used......

8 training finished! average train loss: 0.016367493905803596

total clf loss: 16.59330487110492 total adv loss: 7.135017594427609

evaluating...

Confusion matrix, without normalization

[[ 111 165 18]

[ 118 3626 77]

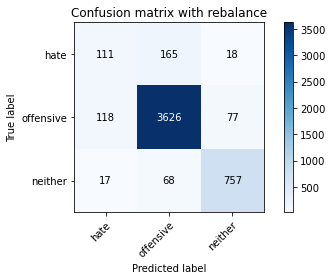
[ 17 68 757]]

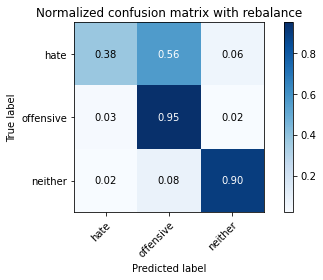
Normalized confusion matrix

[[0.37755102 0.56122449 0.06122449]

[0.03088197 0.94896624 0.02015179]

[0.02019002 0.0807601 0.89904988]]





precision recall f1-score support

hate 0.45 0.38 0.41 294

offensive 0.94 0.95 0.94 3821

neither 0.89 0.90 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: 0.5025296558961864, accuracy: 0.9065967318942909

training 9 epoch...

916.9882354736328 seconds used......

9 training finished! average train loss: 0.013068876952245538

total clf loss: 13.176653131075 total adv loss: 4.246543730941539

evaluating...

Confusion matrix, without normalization

[[ 117 160 17]

[ 144 3601 76]

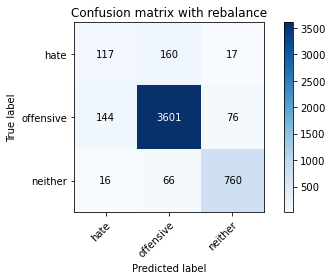
[ 16 66 760]]

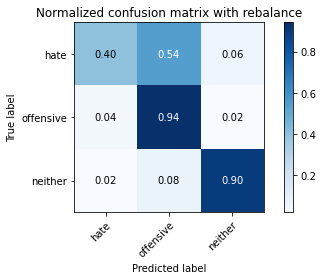
Normalized confusion matrix

[[0.39795918 0.54421769 0.05782313]

[0.03768647 0.94242345 0.01989008]

[0.01900238 0.0783848 0.90261283]]





precision recall f1-score support

hate 0.42 0.40 0.41 294

offensive 0.94 0.94 0.94 3821

neither 0.89 0.90 0.90 842

accuracy 0.90 4957

macro avg 0.75 0.75 0.75 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.5361547934443215, accuracy: 0.9033689731692556

training 10 epoch...

917.3251051902771 seconds used......

10 training finished! average train loss: 0.015886675251089576

total clf loss: 16.15750339353508 total adv loss: 7.958431090648446

evaluating...

Confusion matrix, without normalization

[[ 124 154 16]

[ 160 3589 72]

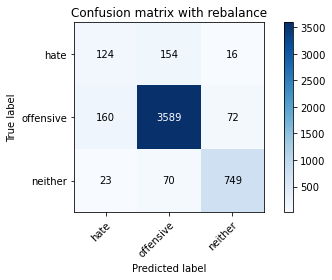
[ 23 70 749]]

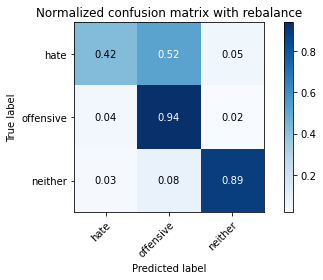
Normalized confusion matrix

[[0.42176871 0.52380952 0.05442177]

[0.04187386 0.93928291 0.01884323]

[0.02731591 0.08313539 0.88954869]]





precision recall f1-score support

hate 0.40 0.42 0.41 294

offensive 0.94 0.94 0.94 3821

neither 0.89 0.89 0.89 842

accuracy 0.90 4957

macro avg 0.75 0.75 0.75 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.4844195652266564, accuracy: 0.9001412144442204

input: last 2 layer

1 run......

Downloading: 100%

411/411 [00:00<00:00, 11.7kB/s]

Downloading: 100%

263M/263M [00:09<00:00, 29.4MB/s]

training 1 epoch...

937.456137418747 seconds used......

1 training finished! average train loss: 0.28406055082732273

total clf loss: 288.7490470148623 total adv loss: 139.21961222816026

evaluating...

Confusion matrix, without normalization

[[ 69 202 23]

[ 54 3693 74]

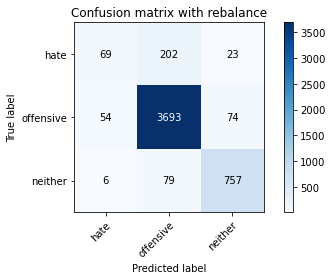
[ 6 79 757]]

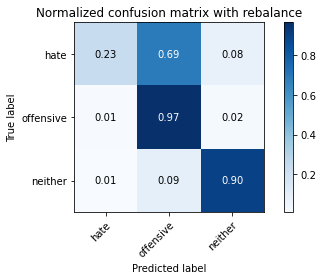
Normalized confusion matrix

[[0.23469388 0.68707483 0.07823129]

[0.01413243 0.96650092 0.01936666]

[0.00712589 0.09382423 0.89904988]]





precision recall f1-score support

hate 0.53 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 0.78 0.70 0.72 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.24676762933602497, accuracy: 0.9116401049021585

training 2 epoch...

939.1131210327148 seconds used......

2 training finished! average train loss: 0.20561333158336986

total clf loss: 208.06117609655485 total adv loss: 81.85501911875326

evaluating...

Confusion matrix, without normalization

[[ 61 216 17]

[ 38 3702 81]

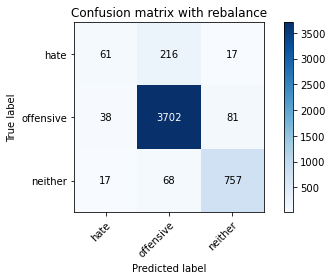
[ 17 68 757]]

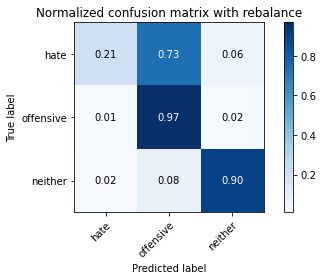
Normalized confusion matrix

[[0.20748299 0.73469388 0.05782313]

[0.00994504 0.96885632 0.02119864]

[0.02019002 0.0807601 0.89904988]]





precision recall f1-score support

hate 0.53 0.21 0.30 294

offensive 0.93 0.97 0.95 3821

neither 0.89 0.90 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.69 0.71 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.24259717969764624, accuracy: 0.9118418398224732

training 3 epoch...

939.3322958946228 seconds used......

3 training finished! average train loss: 0.15516627481303014

total clf loss: 156.42972030118108 total adv loss: 50.0955210046086

evaluating...

Confusion matrix, without normalization

[[ 84 193 17]

[ 74 3660 87]

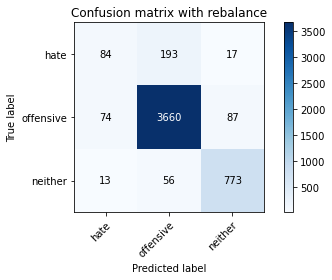
[ 13 56 773]]

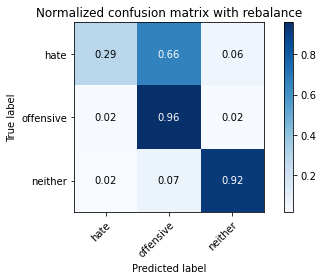
Normalized confusion matrix

[[0.28571429 0.65646259 0.05782313]

[0.01936666 0.95786443 0.02276891]

[0.01543943 0.06650831 0.91805226]]





precision recall f1-score support

hate 0.49 0.29 0.36 294

offensive 0.94 0.96 0.95 3821

neither 0.88 0.92 0.90 842

accuracy 0.91 4957

macro avg 0.77 0.72 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.2775650079326419, accuracy: 0.9112366350615292

training 4 epoch...

938.3081419467926 seconds used......

4 training finished! average train loss: 0.10765600260668678

total clf loss: 108.10971188056283 total adv loss: 26.29914732590987

evaluating...

Confusion matrix, without normalization

[[ 127 148 19]

[ 130 3605 86]

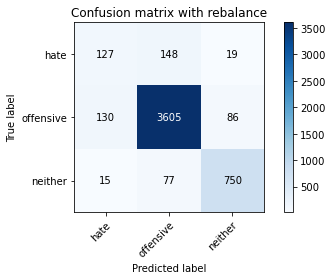
[ 15 77 750]]

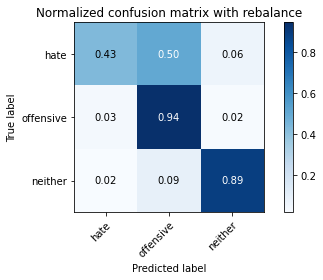
Normalized confusion matrix

[[0.43197279 0.50340136 0.06462585]

[0.03402251 0.9434703 0.0225072 ]

[0.01781473 0.09144893 0.89073634]]





precision recall f1-score support

hate 0.47 0.43 0.45 294

offensive 0.94 0.94 0.94 3821

neither 0.88 0.89 0.88 842

accuracy 0.90 4957

macro avg 0.76 0.76 0.76 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.3187148569146126, accuracy: 0.9041759128505145

training 5 epoch...

938.3944227695465 seconds used......

5 training finished! average train loss: 0.06504418548736852

total clf loss: 65.33745058384375 total adv loss: 16.27237168832653

evaluating...

Confusion matrix, without normalization

[[ 113 163 18]

[ 123 3631 67]

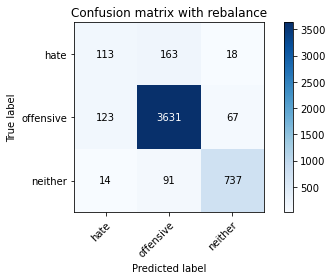
[ 14 91 737]]

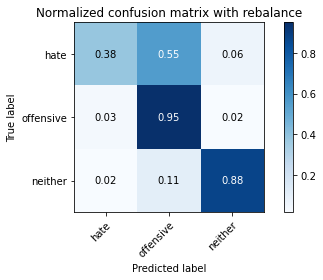
Normalized confusion matrix

[[0.38435374 0.55442177 0.06122449]

[0.03219053 0.9502748 0.01753468]

[0.01662708 0.10807601 0.87529691]]





precision recall f1-score support

hate 0.45 0.38 0.42 294

offensive 0.93 0.95 0.94 3821

neither 0.90 0.88 0.89 842

accuracy 0.90 4957

macro avg 0.76 0.74 0.75 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.36062347399251116, accuracy: 0.9039741779301997

training 6 epoch...

938.5745091438293 seconds used......

6 training finished! average train loss: 0.03589391573805481

total clf loss: 36.17054664035095 total adv loss: 11.275644662881405

evaluating...

Confusion matrix, without normalization

[[ 101 176 17]

[ 114 3630 77]

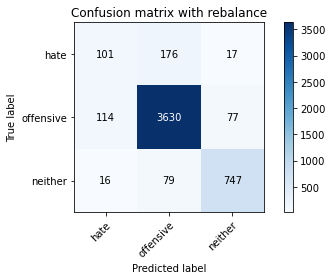
[ 16 79 747]]

Normalized confusion matrix

[[0.34353741 0.59863946 0.05782313]

[0.02983512 0.95001309 0.02015179]

[0.01900238 0.09382423 0.8871734 ]]





precision recall f1-score support

hate 0.44 0.34 0.38 294

offensive 0.93 0.95 0.94 3821

neither 0.89 0.89 0.89 842

accuracy 0.90 4957

macro avg 0.75 0.73 0.74 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.44119896207884, accuracy: 0.9033689731692556

training 7 epoch...

937.5449199676514 seconds used......

7 training finished! average train loss: 0.023628701346128326

total clf loss: 23.993439230885997 total adv loss: 11.075350381760359

evaluating...

Confusion matrix, without normalization

[[ 124 151 19]

[ 140 3605 76]

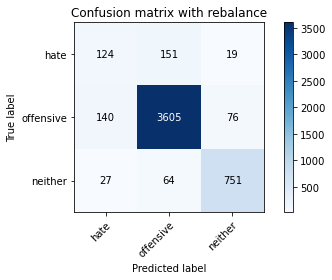
[ 27 64 751]]

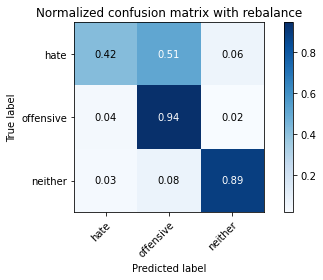
Normalized confusion matrix

[[0.42176871 0.51360544 0.06462585]

[0.03663962 0.9434703 0.01989008]

[0.03206651 0.0760095 0.89192399]]





precision recall f1-score support

hate 0.43 0.42 0.42 294

offensive 0.94 0.94 0.94 3821

neither 0.89 0.89 0.89 842

accuracy 0.90 4957

macro avg 0.75 0.75 0.75 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.47859908495645925, accuracy: 0.903772443009885

training 8 epoch...

937.460489988327 seconds used......

8 training finished! average train loss: 0.019707567859390727

total clf loss: 20.010424867541587 total adv loss: 9.210351747973618

evaluating...

Confusion matrix, without normalization

[[ 142 138 14]

[ 200 3562 59]

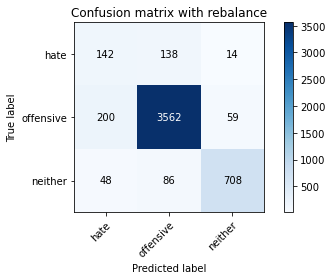
[ 48 86 708]]

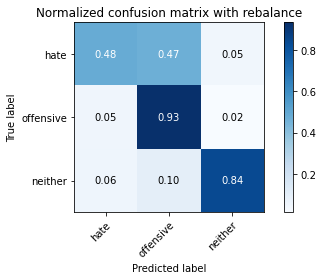
Normalized confusion matrix

[[0.4829932 0.46938776 0.04761905]

[0.05234232 0.9322167 0.01544098]

[0.05700713 0.10213777 0.84085511]]





precision recall f1-score support

hate 0.36 0.48 0.42 294

offensive 0.94 0.93 0.94 3821

neither 0.91 0.84 0.87 842

accuracy 0.89 4957

macro avg 0.74 0.75 0.74 4957

weighted avg 0.90 0.89 0.89 4957

average eval\_loss: 0.5263744342323222, accuracy: 0.890054468428485

training 9 epoch...

937.8677158355713 seconds used......

9 training finished! average train loss: 0.016904300500454477

total clf loss: 17.043992013223033 total adv loss: 5.4985179528355275

evaluating...

Confusion matrix, without normalization

[[ 83 199 12]

[ 87 3691 43]

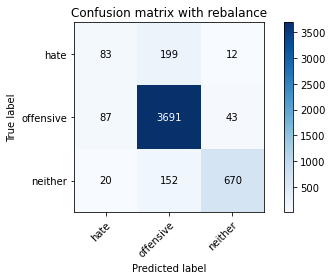
[ 20 152 670]]

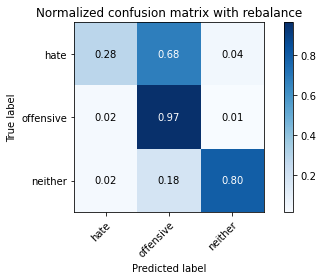
Normalized confusion matrix

[[0.28231293 0.67687075 0.04081633]

[0.02276891 0.96597749 0.0112536 ]

[0.02375297 0.18052257 0.79572447]]





precision recall f1-score support

hate 0.44 0.28 0.34 294

offensive 0.91 0.97 0.94 3821

neither 0.92 0.80 0.86 842

accuracy 0.90 4957

macro avg 0.76 0.68 0.71 4957

weighted avg 0.89 0.90 0.89 4957

average eval\_loss: 0.565601315961225, accuracy: 0.8965099858785556

training 10 epoch...

938.0319299697876 seconds used......

10 training finished! average train loss: 0.015012559786573693

total clf loss: 15.270650570600992 total adv loss: 7.563824905820411

evaluating...

Confusion matrix, without normalization

[[ 117 161 16]

[ 163 3601 57]

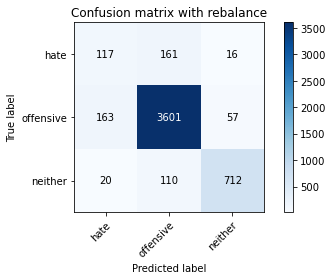
[ 20 110 712]]

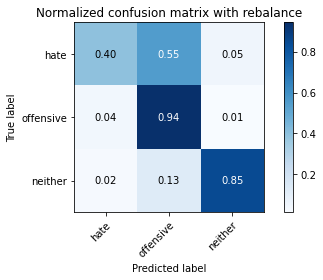
Normalized confusion matrix

[[0.39795918 0.54761905 0.05442177]

[0.04265899 0.94242345 0.01491756]

[0.02375297 0.13064133 0.8456057 ]]





precision recall f1-score support

hate 0.39 0.40 0.39 294

offensive 0.93 0.94 0.94 3821

neither 0.91 0.85 0.88 842

accuracy 0.89 4957

macro avg 0.74 0.73 0.74 4957

weighted avg 0.89 0.89 0.89 4957

average eval\_loss: 0.5338319143571543, accuracy: 0.8936856969941497

2 run......

training 1 epoch...

938.3732113838196 seconds used......

1 training finished! average train loss: 0.27681282053198364

total clf loss: 281.2634081002325 total adv loss: 133.30180232555722

evaluating...

Confusion matrix, without normalization

[[ 92 184 18]

[ 79 3661 81]

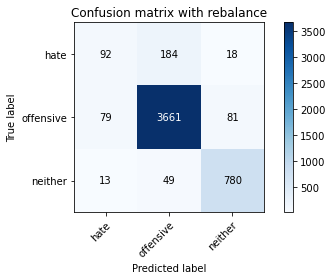
[ 13 49 780]]

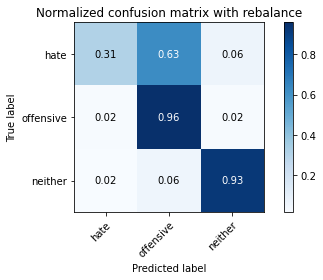
Normalized confusion matrix

[[0.31292517 0.62585034 0.06122449]

[0.02067522 0.95812614 0.02119864]

[0.01543943 0.05819477 0.9263658 ]]





precision recall f1-score support

hate 0.50 0.31 0.38 294

offensive 0.94 0.96 0.95 3821

neither 0.89 0.93 0.91 842

accuracy 0.91 4957

macro avg 0.78 0.73 0.75 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.24335917092168763, accuracy: 0.9144643937865644

training 2 epoch...

939.3617763519287 seconds used......

2 training finished! average train loss: 0.2024383920586608

total clf loss: 204.87723469175398 total adv loss: 81.1669907644391

evaluating...

Confusion matrix, without normalization

[[ 87 180 27]

[ 56 3673 92]

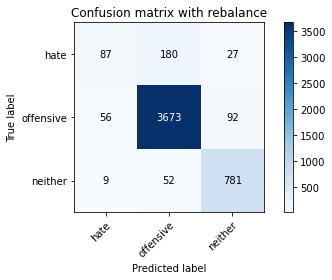
[ 9 52 781]]

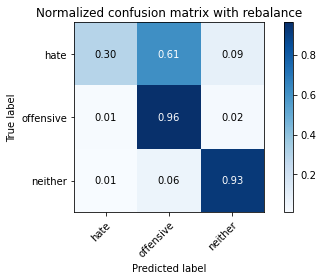
Normalized confusion matrix

[[0.29591837 0.6122449 0.09183673]

[0.01465585 0.96126668 0.02407747]

[0.01068884 0.06175772 0.92755344]]





precision recall f1-score support

hate 0.57 0.30 0.39 294

offensive 0.94 0.96 0.95 3821

neither 0.87 0.93 0.90 842

accuracy 0.92 4957

macro avg 0.79 0.73 0.75 4957

weighted avg 0.91 0.92 0.91 4957

average eval\_loss: 0.24954936337739891, accuracy: 0.9160782731490821

training 3 epoch...

939.5300989151001 seconds used......

3 training finished! average train loss: 0.14440314607007518

total clf loss: 145.59515058528632 total adv loss: 46.94459747581277

evaluating...

Confusion matrix, without normalization

[[ 83 195 16]

[ 58 3690 73]

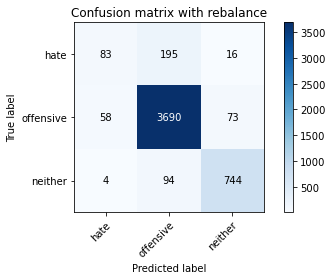
[ 4 94 744]]

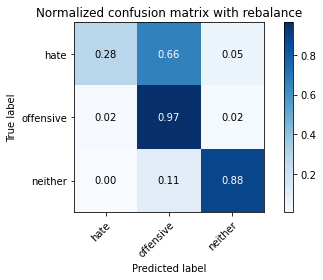
Normalized confusion matrix

[[0.28231293 0.66326531 0.05442177]

[0.01517927 0.96571578 0.01910495]

[0.00475059 0.11163895 0.88361045]]





precision recall f1-score support

hate 0.57 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 0.80 0.71 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.28184613472624354, accuracy: 0.9112366350615292

training 4 epoch...

939.4852683544159 seconds used......

4 training finished! average train loss: 0.09352973900874326

total clf loss: 94.07006110413931 total adv loss: 25.771198992606514

evaluating...

Confusion matrix, without normalization

[[ 119 151 24]

[ 113 3629 79]

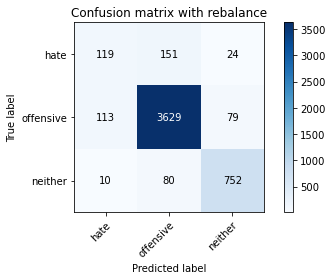
[ 10 80 752]]

Normalized confusion matrix

[[0.4047619 0.51360544 0.08163265]

[0.02957341 0.94975137 0.02067522]

[0.01187648 0.09501188 0.89311164]]





precision recall f1-score support

hate 0.49 0.40 0.44 294

offensive 0.94 0.95 0.94 3821

neither 0.88 0.89 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.75 0.76 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.32049558170594183, accuracy: 0.9078071414161791

training 5 epoch...

939.5466458797455 seconds used......

5 training finished! average train loss: 0.0537446112427255

total clf loss: 54.121478333254345 total adv loss: 16.13647891796427

evaluating...

Confusion matrix, without normalization

[[ 93 185 16]

[ 83 3660 78]

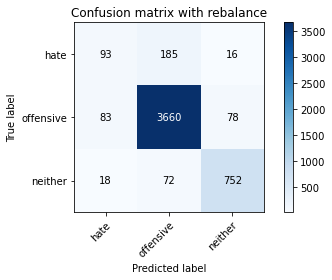
[ 18 72 752]]

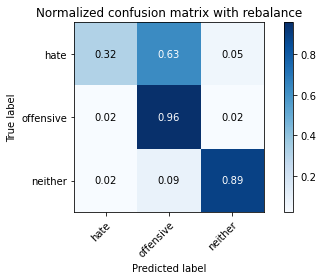
Normalized confusion matrix

[[0.31632653 0.6292517 0.05442177]

[0.02172206 0.95786443 0.0204135 ]

[0.02137767 0.08551069 0.89311164]]





precision recall f1-score support

hate 0.48 0.32 0.38 294

offensive 0.93 0.96 0.95 3821

neither 0.89 0.89 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.72 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.4191197473424367, accuracy: 0.9088158160177526

training 6 epoch...