hate as offensive

alpha=0.05

input: last layer

1 run...

Downloading: 100%

411/411 [00:00<00:00, 14.3kB/s]

Downloading: 100%

263M/263M [00:06<00:00, 46.5MB/s]

training 1 epoch...

115.21826791763306 seconds used......

1 training finished! average train loss: 0.6317268485388561

total clf loss: 110.7531561255455 total adv loss: 54.557309061288834

evaluating...

Confusion matrix, without normalization

[[ 226 36 32]

[ 953 2751 117]

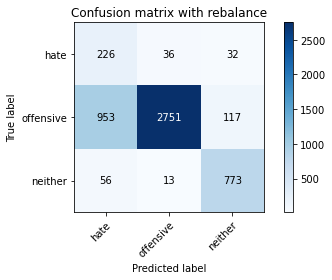
[ 56 13 773]]

Normalized confusion matrix

[[0.76870748 0.12244898 0.10884354]

[0.24941115 0.71996859 0.03062026]

[0.06650831 0.01543943 0.91805226]]





precision recall f1-score support

hate 0.18 0.77 0.30 294

offensive 0.98 0.72 0.83 3821

neither 0.84 0.92 0.88 842

accuracy 0.76 4957

macro avg 0.67 0.80 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 0.5831370730433734, accuracy: 0.7565059511801493

training 2 epoch...

115.18103694915771 seconds used......

2 training finished! average train loss: 0.38851660402894717

total clf loss: 67.72431722283363 total adv loss: 25.75955527368933

evaluating...

Confusion matrix, without normalization

[[ 204 54 36]

[ 578 3068 175]

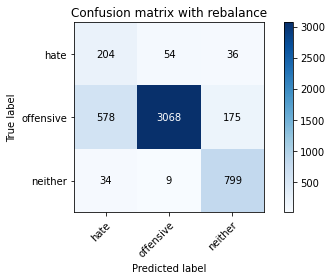
[ 34 9 799]]

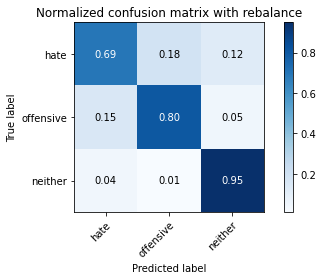
Normalized confusion matrix

[[0.69387755 0.18367347 0.12244898]

[0.1512693 0.80293117 0.04579953]

[0.04038005 0.01068884 0.94893112]]





precision recall f1-score support

hate 0.25 0.69 0.37 294

offensive 0.98 0.80 0.88 3821

neither 0.79 0.95 0.86 842

accuracy 0.82 4957

macro avg 0.67 0.82 0.70 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.46950324138085686, accuracy: 0.82126286060117

training 3 epoch...

115.2500331401825 seconds used......

3 training finished! average train loss: 0.2639667946431372

total clf loss: 45.80572320520878 total adv loss: 13.348024832317606

evaluating...

Confusion matrix, without normalization

[[ 231 46 17]

[ 781 2941 99]

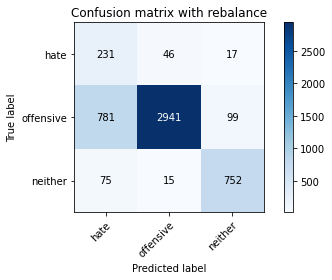
[ 75 15 752]]

Normalized confusion matrix

[[0.78571429 0.15646259 0.05782313]

[0.20439675 0.7696938 0.02590945]

[0.08907363 0.01781473 0.89311164]]





precision recall f1-score support

hate 0.21 0.79 0.33 294

offensive 0.98 0.77 0.86 3821

neither 0.87 0.89 0.88 842

accuracy 0.79 4957

macro avg 0.69 0.82 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.539334298366861, accuracy: 0.7916078273149082

training 4 epoch...

115.25261807441711 seconds used......

4 training finished! average train loss: 0.16461217061382288

total clf loss: 28.402635281905532 total adv loss: 5.079081972769927

evaluating...

Confusion matrix, without normalization

[[ 209 60 25]

[ 707 2988 126]

[ 45 20 777]]

Normalized confusion matrix

[[0.71088435 0.20408163 0.08503401]

[0.1850301 0.78199424 0.03297566]

[0.05344418 0.02375297 0.92280285]]





precision recall f1-score support

hate 0.22 0.71 0.33 294

offensive 0.97 0.78 0.87 3821

neither 0.84 0.92 0.88 842

accuracy 0.80 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.5371063082298685, accuracy: 0.8016945733306435

training 5 epoch...

115.14835953712463 seconds used......

5 training finished! average train loss: 0.10275864494931802

total clf loss: 17.82374514103867 total adv loss: 5.040336202539038

evaluating...

Confusion matrix, without normalization

[[ 230 44 20]

[ 833 2892 96]

[ 64 29 749]]

Normalized confusion matrix

[[0.78231293 0.14965986 0.06802721]

[0.21800576 0.75686993 0.02512431]

[0.0760095 0.03444181 0.88954869]]





precision recall f1-score support

hate 0.20 0.78 0.32 294

offensive 0.98 0.76 0.85 3821

neither 0.87 0.89 0.88 842

accuracy 0.78 4957

macro avg 0.68 0.81 0.68 4957

weighted avg 0.91 0.78 0.83 4957

average eval\_loss: 0.7048024467432932, accuracy: 0.7809158765382288

training 6 epoch...

115.26433777809143 seconds used......

6 training finished! average train loss: 0.05301436121583704

total clf loss: 9.214339616475627 total adv loss: 2.9776773595949635

evaluating...

Confusion matrix, without normalization

[[ 210 66 18]

[ 705 3004 112]

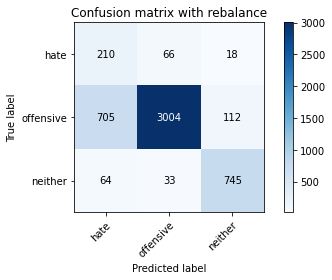
[ 64 33 745]]

Normalized confusion matrix

[[0.71428571 0.2244898 0.06122449]

[0.18450667 0.78618163 0.0293117 ]

[0.0760095 0.0391924 0.8847981 ]]





precision recall f1-score support

hate 0.21 0.71 0.33 294

offensive 0.97 0.79 0.87 3821

neither 0.85 0.88 0.87 842

accuracy 0.80 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.90 0.80 0.84 4957

average eval\_loss: 0.754420843651338, accuracy: 0.7986685495259229

training 7 epoch...

115.26464676856995 seconds used......

7 training finished! average train loss: 0.03586573853213285

total clf loss: 6.191114415414631 total adv loss: 1.1614624396897852

evaluating...

Confusion matrix, without normalization

[[ 218 56 20]

[ 646 3067 108]

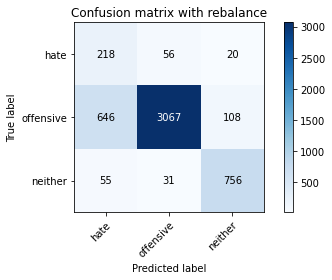
[ 55 31 756]]

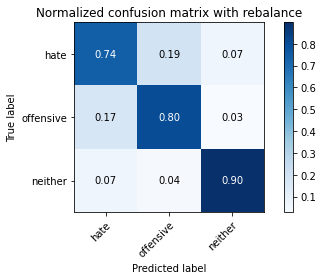
Normalized confusion matrix

[[0.7414966 0.19047619 0.06802721]

[0.16906569 0.80266946 0.02826485]

[0.06532067 0.0368171 0.89786223]]





precision recall f1-score support

hate 0.24 0.74 0.36 294

offensive 0.97 0.80 0.88 3821

neither 0.86 0.90 0.88 842

accuracy 0.82 4957

macro avg 0.69 0.81 0.70 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.8108974495271761, accuracy: 0.8152108129917288

training 8 epoch...

115.24042844772339 seconds used......

8 training finished! average train loss: 0.02527513150609104

total clf loss: 4.450251541449688 total adv loss: 2.5640815075312275

evaluating...

Confusion matrix, without normalization

[[ 225 56 13]

[ 763 2975 83]

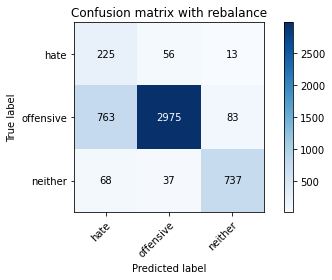
[ 68 37 737]]

Normalized confusion matrix

[[0.76530612 0.19047619 0.04421769]

[0.19968595 0.77859199 0.02172206]

[0.0807601 0.04394299 0.87529691]]





precision recall f1-score support

hate 0.21 0.77 0.33 294

offensive 0.97 0.78 0.86 3821

neither 0.88 0.88 0.88 842

accuracy 0.79 4957

macro avg 0.69 0.81 0.69 4957

weighted avg 0.91 0.79 0.84 4957

average eval\_loss: 1.0062687994326436, accuracy: 0.7942303812789994

training 9 epoch...

115.1515121459961 seconds used......

9 training finished! average train loss: 0.015752150494862643

total clf loss: 2.727687460515881 total adv loss: 0.6813947181217372

evaluating...

Confusion matrix, without normalization

[[ 223 54 17]

[ 759 2948 114]

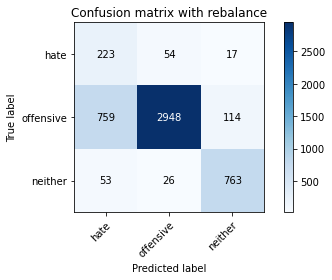
[ 53 26 763]]

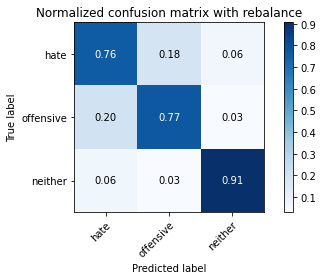
Normalized confusion matrix

[[0.7585034 0.18367347 0.05782313]

[0.1986391 0.77152578 0.02983512]

[0.06294537 0.03087886 0.90617577]]





precision recall f1-score support

hate 0.22 0.76 0.34 294

offensive 0.97 0.77 0.86 3821

neither 0.85 0.91 0.88 842

accuracy 0.79 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 1.200596241851247, accuracy: 0.7936251765180553

training 10 epoch...

115.2439796924591 seconds used......

10 training finished! average train loss: 0.02552360300470296

total clf loss: 4.4597238669521175 total adv loss: 1.9037546510444372

evaluating...

Confusion matrix, without normalization

[[ 169 103 22]

[ 416 3289 116]

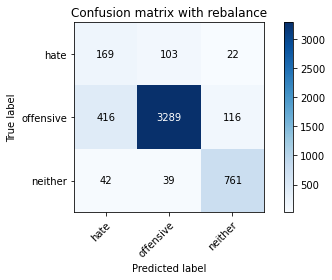
[ 42 39 761]]

Normalized confusion matrix

[[0.57482993 0.35034014 0.07482993]

[0.10887202 0.86076943 0.03035854]

[0.04988124 0.04631829 0.90380048]]





precision recall f1-score support

hate 0.27 0.57 0.37 294

offensive 0.96 0.86 0.91 3821

neither 0.85 0.90 0.87 842

accuracy 0.85 4957

macro avg 0.69 0.78 0.72 4957

weighted avg 0.90 0.85 0.87 4957

average eval\_loss: 0.6879447323679473, accuracy: 0.8511196288077466

2 run...

training 1 epoch...

115.20413398742676 seconds used......

1 training finished! average train loss: 0.6245891769901354

total clf loss: 109.4886282980442 total adv loss: 53.67758156359196

evaluating...

Confusion matrix, without normalization

[[ 183 89 22]

[ 280 3438 103]

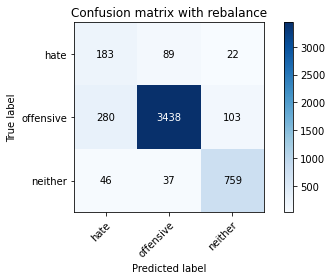
[ 46 37 759]]

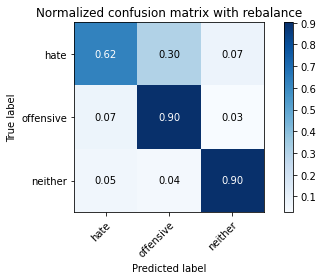
Normalized confusion matrix

[[0.62244898 0.30272109 0.07482993]

[0.07327925 0.89976446 0.02695629]

[0.05463183 0.04394299 0.90142518]]





precision recall f1-score support

hate 0.36 0.62 0.46 294

offensive 0.96 0.90 0.93 3821

neither 0.86 0.90 0.88 842

accuracy 0.88 4957

macro avg 0.73 0.81 0.76 4957

weighted avg 0.91 0.88 0.89 4957

average eval\_loss: 0.35478202575036594, accuracy: 0.8835989509784143

training 2 epoch...

115.24805068969727 seconds used......

2 training finished! average train loss: 0.38204963143631726

total clf loss: 66.61490671336651 total adv loss: 25.688391527161002

evaluating...

Confusion matrix, without normalization

[[ 250 36 8]

[ 876 2868 77]

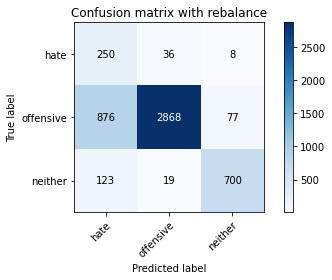
[ 123 19 700]]

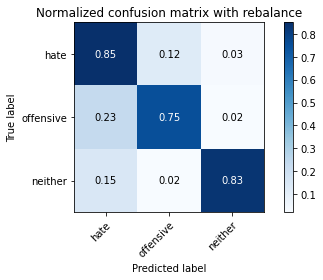
Normalized confusion matrix

[[0.85034014 0.12244898 0.02721088]

[0.22925936 0.75058885 0.02015179]

[0.14608076 0.02256532 0.83135392]]





precision recall f1-score support

hate 0.20 0.85 0.32 294

offensive 0.98 0.75 0.85 3821

neither 0.89 0.83 0.86 842

accuracy 0.77 4957

macro avg 0.69 0.81 0.68 4957

weighted avg 0.92 0.77 0.82 4957

average eval\_loss: 0.5104145702094801, accuracy: 0.7702239257615493

training 3 epoch...

115.24825143814087 seconds used......

3 training finished! average train loss: 0.2633467243428816

total clf loss: 45.7261519767344 total adv loss: 13.877240633126348

evaluating...

Confusion matrix, without normalization

[[ 247 33 14]

[ 973 2759 89]

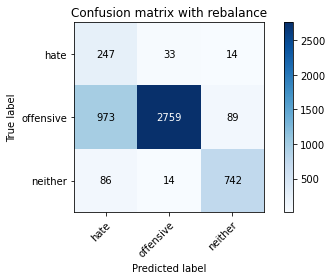
[ 86 14 742]]

Normalized confusion matrix

[[0.84013605 0.1122449 0.04761905]

[0.25464538 0.72206229 0.02329233]

[0.10213777 0.01662708 0.88123515]]





precision recall f1-score support

hate 0.19 0.84 0.31 294

offensive 0.98 0.72 0.83 3821

neither 0.88 0.88 0.88 842

accuracy 0.76 4957

macro avg 0.68 0.81 0.67 4957

weighted avg 0.92 0.76 0.81 4957

average eval\_loss: 0.5642619537369858, accuracy: 0.7561024813395198

training 4 epoch...

115.15428447723389 seconds used......

4 training finished! average train loss: 0.15300413743423963

total clf loss: 26.397087074816227 total adv loss: 4.667592341720592

evaluating...

Confusion matrix, without normalization

[[ 209 59 26]

[ 485 3199 137]

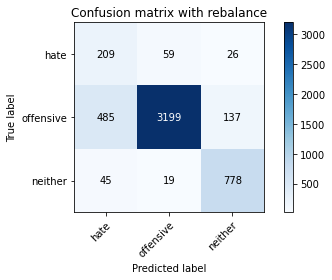
[ 45 19 778]]

Normalized confusion matrix

[[0.71088435 0.20068027 0.08843537]

[0.12693012 0.83721539 0.03585449]

[0.05344418 0.02256532 0.9239905 ]]





precision recall f1-score support

hate 0.28 0.71 0.40 294

offensive 0.98 0.84 0.90 3821

neither 0.83 0.92 0.87 842

accuracy 0.84 4957

macro avg 0.70 0.82 0.73 4957

weighted avg 0.91 0.84 0.87 4957

average eval\_loss: 0.48724281877249237, accuracy: 0.8444623764373613

training 5 epoch...

115.2299485206604 seconds used......

5 training finished! average train loss: 0.10171065112978916

total clf loss: 17.53770472900942 total adv loss: 2.9036676435207482

evaluating...

Confusion matrix, without normalization

[[ 228 51 15]

[ 821 2920 80]

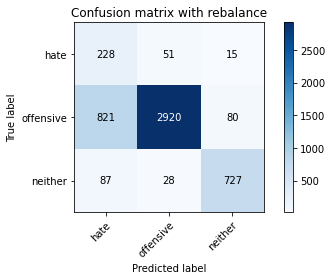
[ 87 28 727]]

Normalized confusion matrix

[[0.7755102 0.17346939 0.05102041]

[0.21486522 0.76419785 0.02093693]

[0.10332542 0.03325416 0.86342043]]





precision recall f1-score support

hate 0.20 0.78 0.32 294

offensive 0.97 0.76 0.86 3821

neither 0.88 0.86 0.87 842

accuracy 0.78 4957

macro avg 0.69 0.80 0.68 4957

weighted avg 0.91 0.78 0.83 4957

average eval\_loss: 0.6943368661788202, accuracy: 0.7817228162194876

training 6 epoch...

115.31119346618652 seconds used......

6 training finished! average train loss: 0.0399320930547533

total clf loss: 6.923385382629931 total adv loss: 1.8999490454880288

evaluating...

Confusion matrix, without normalization

[[ 230 50 14]

[ 752 2983 86]

[ 75 31 736]]

Normalized confusion matrix

[[0.78231293 0.17006803 0.04761905]

[0.19680712 0.78068568 0.0225072 ]

[0.08907363 0.0368171 0.87410926]]





precision recall f1-score support

hate 0.22 0.78 0.34 294

offensive 0.97 0.78 0.87 3821

neither 0.88 0.87 0.88 842

accuracy 0.80 4957

macro avg 0.69 0.81 0.69 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.8473075822281141, accuracy: 0.7966512003227759

training 7 epoch...

115.20353436470032 seconds used......

7 training finished! average train loss: 0.02707132150906503

total clf loss: 4.738718354259618 total adv loss: 2.190447441447759

evaluating...

Confusion matrix, without normalization

[[ 241 40 13]

[ 951 2772 98]

[ 94 16 732]]

Normalized confusion matrix

[[0.81972789 0.13605442 0.04421769]

[0.24888773 0.72546454 0.02564774]

[0.11163895 0.01900238 0.86935867]]





precision recall f1-score support

hate 0.19 0.82 0.31 294

offensive 0.98 0.73 0.83 3821

neither 0.87 0.87 0.87 842

accuracy 0.76 4957

macro avg 0.68 0.80 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 1.1698776847953278, accuracy: 0.7554972765785758

training 8 epoch...

115.01924753189087 seconds used......

8 training finished! average train loss: 0.02151411910791887

total clf loss: 3.7724512877175584 total adv loss: 1.8707386997848516

evaluating...

Confusion matrix, without normalization

[[ 212 59 23]

[ 705 2999 117]

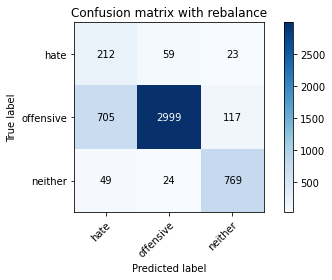
[ 49 24 769]]

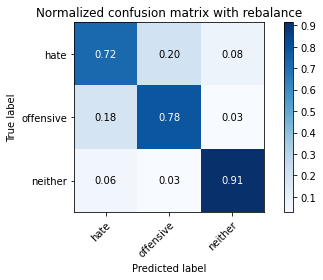
Normalized confusion matrix

[[0.72108844 0.20068027 0.07823129]

[0.18450667 0.78487307 0.03062026]

[0.05819477 0.02850356 0.91330166]]





precision recall f1-score support

hate 0.22 0.72 0.34 294

offensive 0.97 0.78 0.87 3821

neither 0.85 0.91 0.88 842

accuracy 0.80 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.85793572034116, accuracy: 0.8029049828525318

training 9 epoch...

115.11486530303955 seconds used......

9 training finished! average train loss: 0.02134927289018527

total clf loss: 3.8229604471707717 total adv loss: 3.444695576734375

evaluating...

Confusion matrix, without normalization

[[ 167 103 24]

[ 326 3369 126]

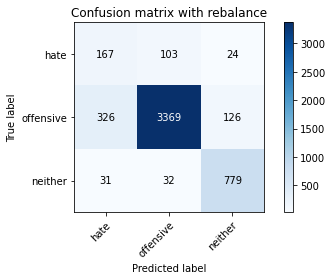
[ 31 32 779]]

Normalized confusion matrix

[[0.56802721 0.35034014 0.08163265]

[0.08531798 0.88170636 0.03297566]

[0.0368171 0.03800475 0.92517815]]





precision recall f1-score support

hate 0.32 0.57 0.41 294

offensive 0.96 0.88 0.92 3821

neither 0.84 0.93 0.88 842

accuracy 0.87 4957

macro avg 0.71 0.79 0.74 4957

weighted avg 0.90 0.87 0.88 4957

average eval\_loss: 0.6536020642642172, accuracy: 0.8704861811579584

training 10 epoch...

114.98825311660767 seconds used......

10 training finished! average train loss: 0.018580369853257332

total clf loss: 3.308355188317364 total adv loss: 2.6222393179195933

evaluating...

Confusion matrix, without normalization

[[ 242 37 15]

[ 944 2769 108]

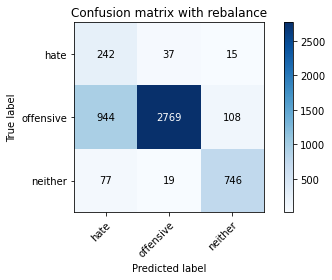
[ 77 19 746]]

Normalized confusion matrix

[[0.82312925 0.12585034 0.05102041]

[0.24705574 0.7246794 0.02826485]

[0.09144893 0.02256532 0.88598575]]





precision recall f1-score support

hate 0.19 0.82 0.31 294

offensive 0.98 0.72 0.83 3821

neither 0.86 0.89 0.87 842

accuracy 0.76 4957

macro avg 0.68 0.81 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 1.4013961992722246, accuracy: 0.7579180956223522

3 run...

training 1 epoch...

114.96818089485168 seconds used......

1 training finished! average train loss: 0.6393965381627892

total clf loss: 112.01004299521446 total adv loss: 53.464704208076

evaluating...

Confusion matrix, without normalization

[[ 209 45 40]

[ 686 2940 195]

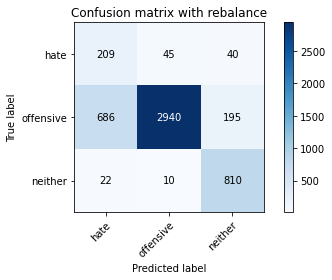
[ 22 10 810]]

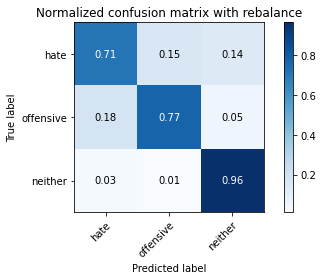
Normalized confusion matrix

[[0.71088435 0.15306122 0.13605442]

[0.17953415 0.76943209 0.05103376]

[0.02612827 0.01187648 0.96199525]]





precision recall f1-score support

hate 0.23 0.71 0.35 294

offensive 0.98 0.77 0.86 3821

neither 0.78 0.96 0.86 842

accuracy 0.80 4957

macro avg 0.66 0.81 0.69 4957

weighted avg 0.90 0.80 0.83 4957

average eval\_loss: 0.504699912943667, accuracy: 0.7986685495259229

training 2 epoch...

115.0699691772461 seconds used......

2 training finished! average train loss: 0.3894779110076832

total clf loss: 67.97908011078835 total adv loss: 27.56714512128383

evaluating...

Confusion matrix, without normalization

[[ 239 42 13]

[ 792 2952 77]

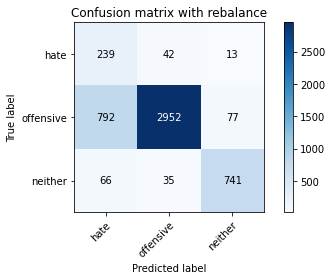
[ 66 35 741]]

Normalized confusion matrix

[[0.81292517 0.14285714 0.04421769]

[0.20727558 0.77257262 0.02015179]

[0.0783848 0.0415677 0.88004751]]





precision recall f1-score support

hate 0.22 0.81 0.34 294

offensive 0.97 0.77 0.86 3821

neither 0.89 0.88 0.89 842

accuracy 0.79 4957

macro avg 0.69 0.82 0.70 4957

weighted avg 0.92 0.79 0.84 4957

average eval\_loss: 0.45066438723475705, accuracy: 0.7932217066774259

training 3 epoch...

114.98005700111389 seconds used......

3 training finished! average train loss: 0.26387903392750617

total clf loss: 45.909043837338686 total adv loss: 15.7145770881325

evaluating...

Confusion matrix, without normalization

[[ 242 33 19]

[ 989 2708 124]

[ 49 13 780]]

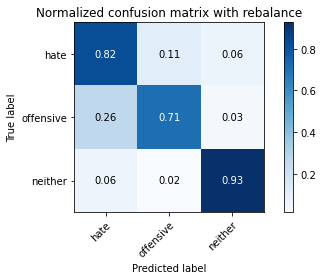
Normalized confusion matrix

[[0.82312925 0.1122449 0.06462585]

[0.25883277 0.708715 0.03245224]

[0.05819477 0.01543943 0.9263658 ]]





precision recall f1-score support

hate 0.19 0.82 0.31 294

offensive 0.98 0.71 0.82 3821

neither 0.85 0.93 0.88 842

accuracy 0.75 4957

macro avg 0.67 0.82 0.67 4957

weighted avg 0.91 0.75 0.80 4957

average eval\_loss: 0.6451664144293435, accuracy: 0.7524712527738552

training 4 epoch...

115.01489853858948 seconds used......

4 training finished! average train loss: 0.15741779962935817

total clf loss: 27.183788870926946 total adv loss: 5.30690301960567

evaluating...

Confusion matrix, without normalization

[[ 225 55 14]

[ 685 3072 64]

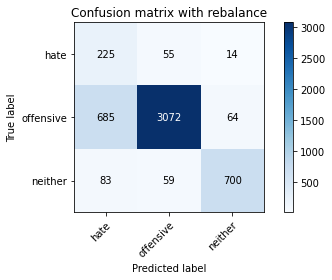
[ 83 59 700]]

Normalized confusion matrix

[[0.76530612 0.18707483 0.04761905]

[0.17927244 0.80397802 0.01674954]

[0.09857482 0.07007126 0.83135392]]





precision recall f1-score support

hate 0.23 0.77 0.35 294

offensive 0.96 0.80 0.88 3821

neither 0.90 0.83 0.86 842

accuracy 0.81 4957

macro avg 0.70 0.80 0.70 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 0.5225778533175828, accuracy: 0.8063344764978818

training 5 epoch...

114.98751163482666 seconds used......

5 training finished! average train loss: 0.07791213071442138

total clf loss: 13.467121494468302 total adv loss: 2.882942590425955

evaluating...

Confusion matrix, without normalization

[[ 215 62 17]

[ 603 3090 128]

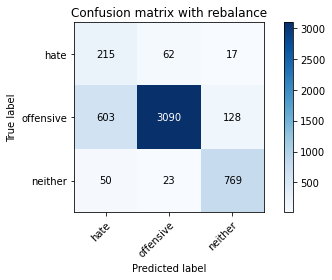
[ 50 23 769]]

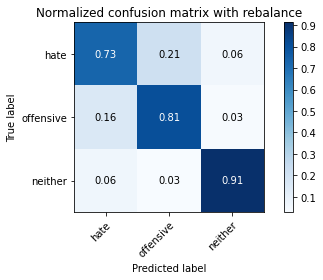
Normalized confusion matrix

[[0.73129252 0.21088435 0.05782313]

[0.15781209 0.80868882 0.03349908]

[0.05938242 0.02731591 0.91330166]]





precision recall f1-score support

hate 0.25 0.73 0.37 294

offensive 0.97 0.81 0.88 3821

neither 0.84 0.91 0.88 842

accuracy 0.82 4957

macro avg 0.69 0.82 0.71 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.6021551416976557, accuracy: 0.8218680653621142

training 6 epoch...

115.11857891082764 seconds used......

6 training finished! average train loss: 0.055007800901695346

total clf loss: 9.541117830900475 total adv loss: 2.695677906900528

evaluating...

Confusion matrix, without normalization

[[ 222 48 24]

[ 815 2886 120]

[ 50 30 762]]

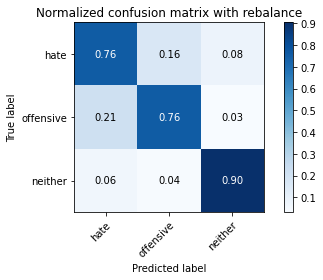
Normalized confusion matrix

[[0.75510204 0.16326531 0.08163265]

[0.21329495 0.75529966 0.03140539]

[0.05938242 0.03562945 0.90498812]]





precision recall f1-score support

hate 0.20 0.76 0.32 294

offensive 0.97 0.76 0.85 3821

neither 0.84 0.90 0.87 842

accuracy 0.78 4957

macro avg 0.67 0.81 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.877258341338846, accuracy: 0.7807141416179141

training 7 epoch...

115.05860757827759 seconds used......

7 training finished! average train loss: 0.02984585040841219

total clf loss: 5.165848835546058 total adv loss: 1.2441679269250017

evaluating...

Confusion matrix, without normalization

[[ 229 49 16]

[ 779 2947 95]

[ 54 44 744]]

Normalized confusion matrix

[[0.77891156 0.16666667 0.05442177]

[0.20387333 0.77126407 0.0248626 ]

[0.06413302 0.05225653 0.88361045]]





precision recall f1-score support

hate 0.22 0.78 0.34 294

offensive 0.97 0.77 0.86 3821

neither 0.87 0.88 0.88 842

accuracy 0.79 4957

macro avg 0.69 0.81 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.8759345897931545, accuracy: 0.7908008876336494

training 8 epoch...

115.00230526924133 seconds used......

8 training finished! average train loss: 0.019995840830722967

total clf loss: 3.4717187136993743 total adv loss: 1.0485984958213521

evaluating...

Confusion matrix, without normalization

[[ 207 72 15]

[ 553 3179 89]

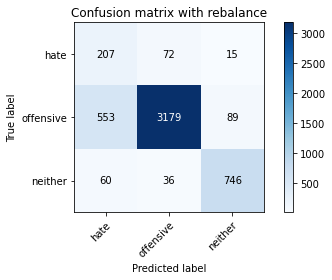
[ 60 36 746]]

Normalized confusion matrix

[[0.70408163 0.24489796 0.05102041]

[0.14472651 0.83198116 0.02329233]

[0.07125891 0.04275534 0.88598575]]





precision recall f1-score support

hate 0.25 0.70 0.37 294

offensive 0.97 0.83 0.89 3821

neither 0.88 0.89 0.88 842

accuracy 0.83 4957

macro avg 0.70 0.81 0.72 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.7980050306923071, accuracy: 0.8335686907403672

training 9 epoch...

115.07380199432373 seconds used......

9 training finished! average train loss: 0.011990194926133695

total clf loss: 2.108123964368133 total adv loss: 1.1560123684394057

evaluating...

Confusion matrix, without normalization

[[ 199 72 23]

[ 453 3257 111]

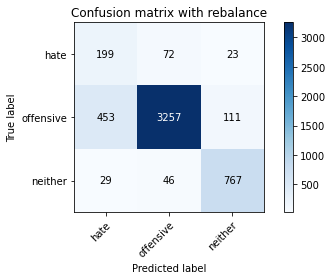
[ 29 46 767]]

Normalized confusion matrix

[[0.67687075 0.24489796 0.07823129]

[0.11855535 0.85239466 0.02904999]

[0.03444181 0.05463183 0.91092637]]





precision recall f1-score support

hate 0.29 0.68 0.41 294

offensive 0.97 0.85 0.91 3821

neither 0.85 0.91 0.88 842

accuracy 0.85 4957

macro avg 0.70 0.81 0.73 4957

weighted avg 0.91 0.85 0.87 4957

average eval\_loss: 0.7708812929983739, accuracy: 0.8519265684890055

training 10 epoch...

115.05061531066895 seconds used......

10 training finished! average train loss: 0.01423822206774201

total clf loss: 2.4943148134625517 total adv loss: 1.1915765751036815

evaluating...

Confusion matrix, without normalization

[[ 226 51 17]

[ 873 2848 100]

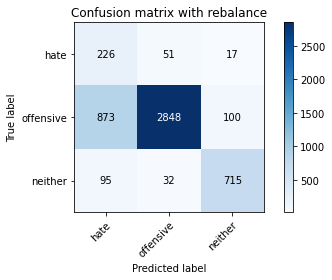
[ 95 32 715]]

Normalized confusion matrix

[[0.76870748 0.17346939 0.05782313]

[0.22847422 0.74535462 0.02617116]

[0.1128266 0.03800475 0.84916865]]





precision recall f1-score support

hate 0.19 0.77 0.30 294

offensive 0.97 0.75 0.84 3821

neither 0.86 0.85 0.85 842

accuracy 0.76 4957

macro avg 0.67 0.79 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 1.2911231958066984, accuracy: 0.7643736130724228

4 run...

training 1 epoch...

115.1001935005188 seconds used......

1 training finished! average train loss: 0.6467839693813994

total clf loss: 113.35433252155781 total adv loss: 55.08547543361783

evaluating...

Confusion matrix, without normalization

[[ 195 70 29]

[ 406 3282 133]

[ 48 18 776]]

Normalized confusion matrix

[[0.66326531 0.23809524 0.09863946]

[0.10625491 0.85893745 0.03480764]

[0.05700713 0.02137767 0.9216152 ]]





precision recall f1-score support

hate 0.30 0.66 0.41 294

offensive 0.97 0.86 0.91 3821

neither 0.83 0.92 0.87 842

accuracy 0.86 4957

macro avg 0.70 0.81 0.73 4957

weighted avg 0.91 0.86 0.88 4957

average eval\_loss: 0.4003343879575691, accuracy: 0.8579786160984466

training 2 epoch...

115.31490778923035 seconds used......

2 training finished! average train loss: 0.3918803630516543

total clf loss: 68.45998068153858 total adv loss: 28.968775475397706

evaluating...

Confusion matrix, without normalization

[[ 227 36 31]

[ 806 2868 147]

[ 28 14 800]]

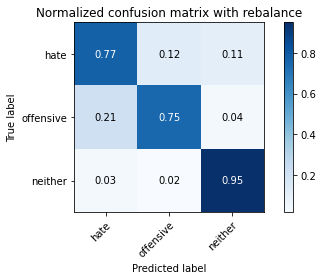
Normalized confusion matrix

[[0.77210884 0.12244898 0.10544218]

[0.21093954 0.75058885 0.0384716 ]

[0.03325416 0.01662708 0.95011876]]





precision recall f1-score support

hate 0.21 0.77 0.34 294

offensive 0.98 0.75 0.85 3821

neither 0.82 0.95 0.88 842

accuracy 0.79 4957

macro avg 0.67 0.82 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.4907508595816551, accuracy: 0.7857575146257817

training 3 epoch...

115.09898924827576 seconds used......

3 training finished! average train loss: 0.2775373582167235

total clf loss: 48.223484095185995 total adv loss: 15.291915968758985

evaluating...

Confusion matrix, without normalization

[[ 225 55 14]

[ 613 3119 89]

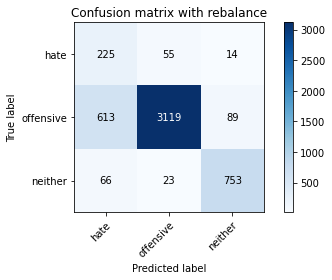
[ 66 23 753]]

Normalized confusion matrix

[[0.76530612 0.18707483 0.04761905]

[0.16042921 0.81627846 0.02329233]

[0.0783848 0.02731591 0.89429929]]





precision recall f1-score support

hate 0.25 0.77 0.38 294

offensive 0.98 0.82 0.89 3821

neither 0.88 0.89 0.89 842

accuracy 0.83 4957

macro avg 0.70 0.83 0.72 4957

weighted avg 0.92 0.83 0.86 4957

average eval\_loss: 0.4382406806483144, accuracy: 0.8265079685293524

training 4 epoch...

115.13669276237488 seconds used......

4 training finished! average train loss: 0.17671614361393173

total clf loss: 30.52927527949214 total adv loss: 6.21629327267874

evaluating...

Confusion matrix, without normalization

[[ 221 58 15]

[ 682 3044 95]

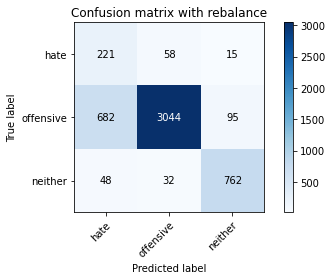
[ 48 32 762]]

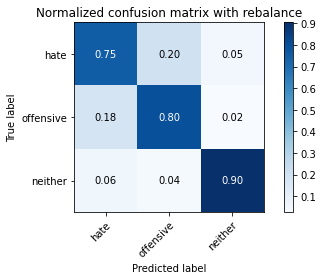
Normalized confusion matrix

[[0.75170068 0.19727891 0.05102041]

[0.17848731 0.79665009 0.0248626 ]

[0.05700713 0.03800475 0.90498812]]





precision recall f1-score support

hate 0.23 0.75 0.36 294

offensive 0.97 0.80 0.88 3821

neither 0.87 0.90 0.89 842

accuracy 0.81 4957

macro avg 0.69 0.82 0.71 4957

weighted avg 0.91 0.81 0.85 4957

average eval\_loss: 0.4823844043477889, accuracy: 0.8123865241073229

training 5 epoch...

115.14335179328918 seconds used......

5 training finished! average train loss: 0.09326464705147415

total clf loss: 16.07922767754644 total adv loss: 2.6194614464911865

evaluating...

Confusion matrix, without normalization

[[ 212 59 23]

[ 651 3055 115]

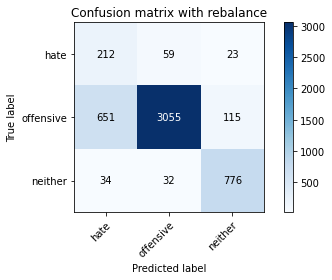
[ 34 32 776]]

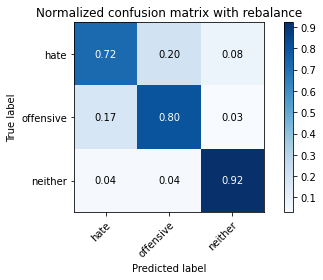
Normalized confusion matrix

[[0.72108844 0.20068027 0.07823129]

[0.17037425 0.79952892 0.03009683]

[0.04038005 0.03800475 0.9216152 ]]





precision recall f1-score support

hate 0.24 0.72 0.36 294

offensive 0.97 0.80 0.88 3821

neither 0.85 0.92 0.88 842

accuracy 0.82 4957

macro avg 0.69 0.81 0.71 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.6206011777187908, accuracy: 0.8156142828323583

training 6 epoch...

115.28244757652283 seconds used......

6 training finished! average train loss: 0.048680567796674304

total clf loss: 8.426022456260398 total adv loss: 2.0329071383021073

evaluating...

Confusion matrix, without normalization

[[ 205 78 11]

[ 517 3219 85]

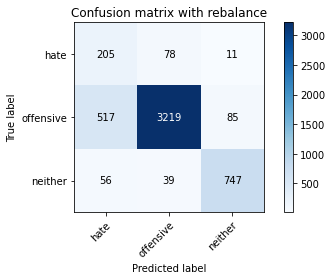
[ 56 39 747]]

Normalized confusion matrix

[[0.69727891 0.26530612 0.03741497]

[0.13530489 0.84244962 0.02224549]

[0.06650831 0.04631829 0.8871734 ]]





precision recall f1-score support

hate 0.26 0.70 0.38 294

offensive 0.96 0.84 0.90 3821

neither 0.89 0.89 0.89 842

accuracy 0.84 4957

macro avg 0.70 0.81 0.72 4957

weighted avg 0.91 0.84 0.87 4957

average eval\_loss: 0.6731686680896899, accuracy: 0.8414363526326407

training 7 epoch...

115.24494361877441 seconds used......

7 training finished! average train loss: 0.03746242980385602

total clf loss: 6.461453223135322 total adv loss: 1.1075543917540926

evaluating...

Confusion matrix, without normalization

[[ 219 58 17]

[ 773 2927 121]

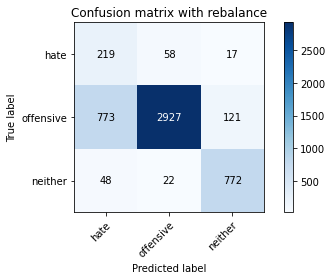
[ 48 22 772]]

Normalized confusion matrix

[[0.74489796 0.19727891 0.05782313]

[0.20230306 0.76602984 0.0316671 ]

[0.05700713 0.02612827 0.91686461]]





precision recall f1-score support

hate 0.21 0.74 0.33 294

offensive 0.97 0.77 0.86 3821

neither 0.85 0.92 0.88 842

accuracy 0.79 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.8895618470626012, accuracy: 0.79039741779302

training 8 epoch...

115.32541370391846 seconds used......

8 training finished! average train loss: 0.030066258916345408

total clf loss: 5.184571422520094 total adv loss: 0.8648231931947521

evaluating...

Confusion matrix, without normalization

[[ 221 54 19]

[ 777 2925 119]

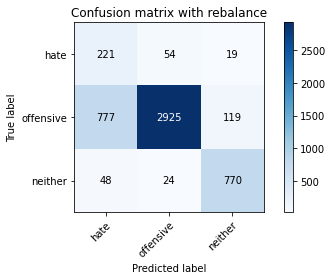
[ 48 24 770]]

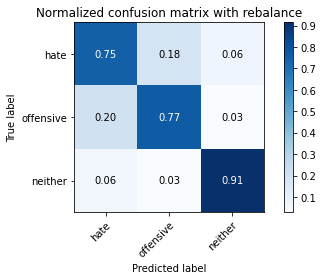
Normalized confusion matrix

[[0.75170068 0.18367347 0.06462585]

[0.20334991 0.76550641 0.03114368]

[0.05700713 0.02850356 0.91448931]]





precision recall f1-score support

hate 0.21 0.75 0.33 294

offensive 0.97 0.77 0.86 3821

neither 0.85 0.91 0.88 842

accuracy 0.79 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 1.0146969470041503, accuracy: 0.7899939479523905

training 9 epoch...

115.17600512504578 seconds used......

9 training finished! average train loss: 0.016174872518498015

total clf loss: 2.8332326858653687 total adv loss: 1.3465898531649145

evaluating...

Confusion matrix, without normalization

[[ 190 74 30]

[ 474 3210 137]

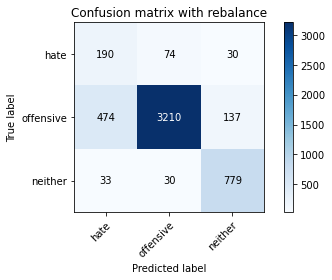
[ 33 30 779]]

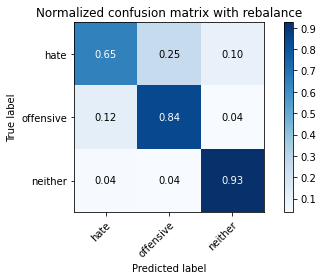
Normalized confusion matrix

[[0.6462585 0.25170068 0.10204082]

[0.1240513 0.84009422 0.03585449]

[0.0391924 0.03562945 0.92517815]]





precision recall f1-score support

hate 0.27 0.65 0.38 294

offensive 0.97 0.84 0.90 3821

neither 0.82 0.93 0.87 842

accuracy 0.84 4957

macro avg 0.69 0.80 0.72 4957

weighted avg 0.90 0.84 0.86 4957

average eval\_loss: 0.7946061734855315, accuracy: 0.8430502319951584

training 10 epoch...

115.20471024513245 seconds used......

10 training finished! average train loss: 0.014065047864235763

total clf loss: 2.4324728891660925 total adv loss: 0.5469939788163174

evaluating...

Confusion matrix, without normalization

[[ 213 59 22]

[ 767 2935 119]

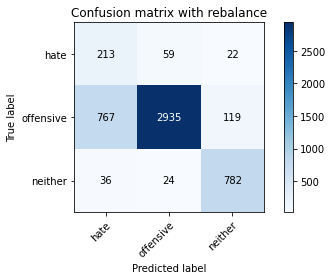
[ 36 24 782]]

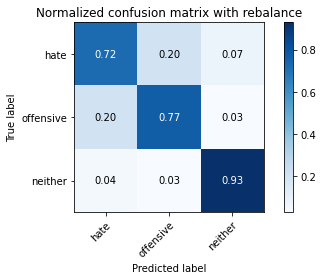
Normalized confusion matrix

[[0.7244898 0.20068027 0.07482993]

[0.20073279 0.76812353 0.03114368]

[0.04275534 0.02850356 0.92874109]]





precision recall f1-score support

hate 0.21 0.72 0.33 294

offensive 0.97 0.77 0.86 3821

neither 0.85 0.93 0.89 842

accuracy 0.79 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 1.0924408520335303, accuracy: 0.7928182368367964

5 run...

training 1 epoch...

115.27380323410034 seconds used......

1 training finished! average train loss: 0.6474021611332196

total clf loss: 113.3634218275547 total adv loss: 53.15304550155997

evaluating...

Confusion matrix, without normalization

[[ 244 37 13]

[ 890 2847 84]

[ 115 15 712]]

Normalized confusion matrix

[[0.82993197 0.12585034 0.04421769]

[0.23292332 0.74509291 0.02198377]

[0.13657957 0.01781473 0.8456057 ]]





precision recall f1-score support

hate 0.20 0.83 0.32 294

offensive 0.98 0.75 0.85 3821

neither 0.88 0.85 0.86 842

accuracy 0.77 4957

macro avg 0.69 0.81 0.68 4957

weighted avg 0.92 0.77 0.82 4957

average eval\_loss: 0.5146386088623155, accuracy: 0.7671979019568287

training 2 epoch...

115.21164846420288 seconds used......

2 training finished! average train loss: 0.3951620686559649

total clf loss: 68.94329833984375 total adv loss: 27.411691970191896

evaluating...

Confusion matrix, without normalization

[[ 219 53 22]

[ 663 3051 107]

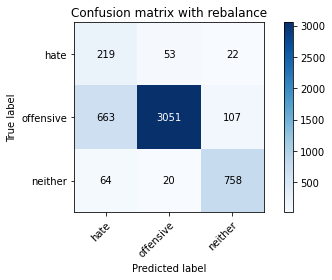
[ 64 20 758]]

Normalized confusion matrix

[[0.74489796 0.18027211 0.07482993]

[0.17351479 0.79848207 0.02800314]

[0.0760095 0.02375297 0.90023753]]





precision recall f1-score support

hate 0.23 0.74 0.35 294

offensive 0.98 0.80 0.88 3821

neither 0.85 0.90 0.88 842

accuracy 0.81 4957

macro avg 0.69 0.81 0.70 4957

weighted avg 0.91 0.81 0.85 4957

average eval\_loss: 0.43144314769174785, accuracy: 0.8125882590276376

training 3 epoch...

115.16143822669983 seconds used......

3 training finished! average train loss: 0.2746397747020972

total clf loss: 47.65579389035702 total adv loss: 13.847846856573597

evaluating...

Confusion matrix, without normalization

[[ 214 55 25]

[ 576 3107 138]

[ 42 20 780]]

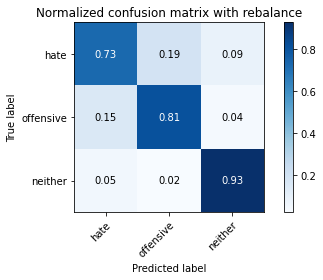
Normalized confusion matrix

[[0.72789116 0.18707483 0.08503401]

[0.15074588 0.81313792 0.0361162 ]

[0.04988124 0.02375297 0.9263658 ]]





precision recall f1-score support

hate 0.26 0.73 0.38 294

offensive 0.98 0.81 0.89 3821

neither 0.83 0.93 0.87 842

accuracy 0.83 4957

macro avg 0.69 0.82 0.71 4957

weighted avg 0.91 0.83 0.85 4957

average eval\_loss: 0.44945593254129973, accuracy: 0.8273149082106113

training 4 epoch...

115.20936298370361 seconds used......

4 training finished! average train loss: 0.17042297886259722

total clf loss: 29.490063020028174 total adv loss: 6.95467277726857

evaluating...

Confusion matrix, without normalization

[[ 229 42 23]

[ 955 2702 164]

[ 52 13 777]]

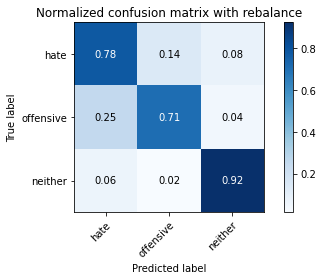
Normalized confusion matrix

[[0.77891156 0.14285714 0.07823129]

[0.24993457 0.70714473 0.0429207 ]

[0.06175772 0.01543943 0.92280285]]





precision recall f1-score support

hate 0.19 0.78 0.30 294

offensive 0.98 0.71 0.82 3821

neither 0.81 0.92 0.86 842

accuracy 0.75 4957

macro avg 0.66 0.80 0.66 4957

weighted avg 0.90 0.75 0.80 4957

average eval\_loss: 0.7712946123473586, accuracy: 0.7480330845269316

training 5 epoch...

115.13847208023071 seconds used......

5 training finished! average train loss: 0.09673852211801194

total clf loss: 16.69668514560908 total adv loss: 3.08795702556381

evaluating...

Confusion matrix, without normalization

[[ 215 59 20]

[ 558 3118 145]

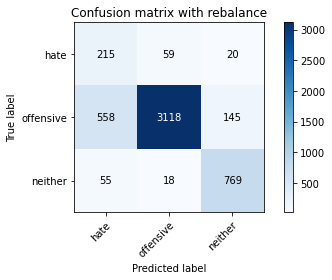
[ 55 18 769]]

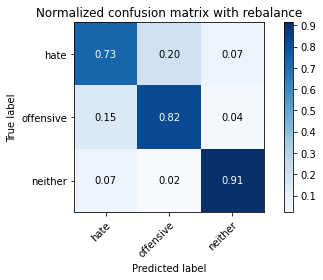
Normalized confusion matrix

[[0.73129252 0.20068027 0.06802721]

[0.14603507 0.81601675 0.03794818]

[0.06532067 0.02137767 0.91330166]]





precision recall f1-score support

hate 0.26 0.73 0.38 294

offensive 0.98 0.82 0.89 3821

neither 0.82 0.91 0.87 842

accuracy 0.83 4957

macro avg 0.69 0.82 0.71 4957

weighted avg 0.91 0.83 0.85 4957

average eval\_loss: 0.6706275671932681, accuracy: 0.827516643130926

training 6 epoch...

115.23362994194031 seconds used......

6 training finished! average train loss: 0.05272001416520461

total clf loss: 9.13544743671082 total adv loss: 2.406500968616456

evaluating...

Confusion matrix, without normalization

[[ 212 63 19]

[ 522 3194 105]

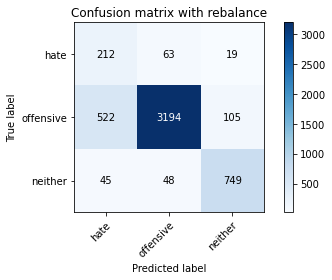
[ 45 48 749]]

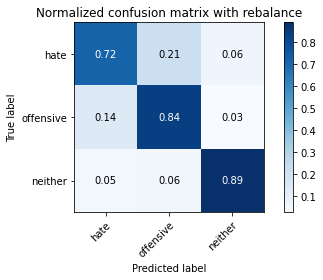
Normalized confusion matrix

[[0.72108844 0.21428571 0.06462585]

[0.13661345 0.83590683 0.02747972]

[0.05344418 0.05700713 0.88954869]]





precision recall f1-score support

hate 0.27 0.72 0.40 294

offensive 0.97 0.84 0.90 3821

neither 0.86 0.89 0.87 842

accuracy 0.84 4957

macro avg 0.70 0.82 0.72 4957

weighted avg 0.91 0.84 0.86 4957

average eval\_loss: 0.6671455247551128, accuracy: 0.8382085939076054

training 7 epoch...

115.09825325012207 seconds used......

7 training finished! average train loss: 0.031720169677946455

total clf loss: 5.48086302774027 total adv loss: 1.1342801449209219

evaluating...

Confusion matrix, without normalization

[[ 231 46 17]

[ 841 2871 109]

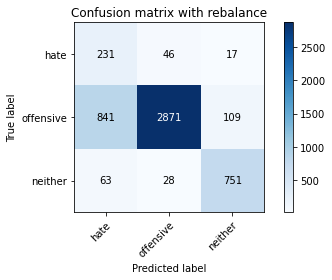
[ 63 28 751]]

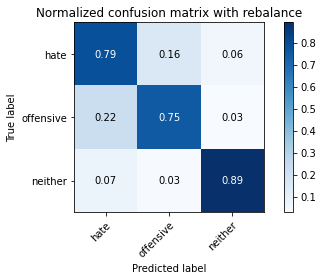
Normalized confusion matrix

[[0.78571429 0.15646259 0.05782313]

[0.22009945 0.75137399 0.02852656]

[0.07482185 0.03325416 0.89192399]]





precision recall f1-score support

hate 0.20 0.79 0.32 294

offensive 0.97 0.75 0.85 3821

neither 0.86 0.89 0.87 842

accuracy 0.78 4957

macro avg 0.68 0.81 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 1.0032756170333033, accuracy: 0.777284647972564

training 8 epoch...

115.23242688179016 seconds used......

8 training finished! average train loss: 0.029857295453997755

total clf loss: 5.171379058738239 total adv loss: 1.3156308756151702

evaluating...

Confusion matrix, without normalization

[[ 239 42 13]

[ 965 2758 98]

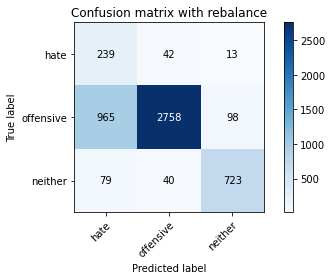
[ 79 40 723]]

Normalized confusion matrix

[[0.81292517 0.14285714 0.04421769]

[0.25255169 0.72180058 0.02564774]

[0.09382423 0.04750594 0.85866983]]





precision recall f1-score support

hate 0.19 0.81 0.30 294

offensive 0.97 0.72 0.83 3821

neither 0.87 0.86 0.86 842

accuracy 0.75 4957

macro avg 0.67 0.80 0.66 4957

weighted avg 0.91 0.75 0.80 4957

average eval\_loss: 1.116007431252529, accuracy: 0.7504539035707081

training 9 epoch...

115.09615588188171 seconds used......

9 training finished! average train loss: 0.023696990298346904

total clf loss: 4.0961329457350075 total adv loss: 0.8789525832107756

evaluating...

Confusion matrix, without normalization

[[ 224 53 17]

[ 708 2986 127]

[ 63 29 750]]

Normalized confusion matrix

[[0.76190476 0.18027211 0.05782313]

[0.18529181 0.78147082 0.03323737]

[0.07482185 0.03444181 0.89073634]]





precision recall f1-score support

hate 0.23 0.76 0.35 294

offensive 0.97 0.78 0.87 3821

neither 0.84 0.89 0.86 842

accuracy 0.80 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 1.0068613475173591, accuracy: 0.7988702844462376

training 10 epoch...

114.9967393875122 seconds used......

10 training finished! average train loss: 0.007427145778208328

total clf loss: 1.3042061159212608 total adv loss: 0.6832834732667834

evaluating...

Confusion matrix, without normalization

[[ 209 66 19]

[ 485 3234 102]

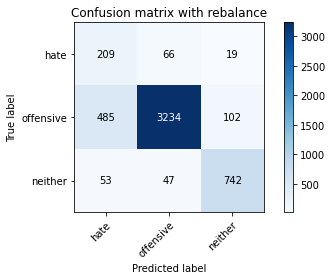
[ 53 47 742]]

Normalized confusion matrix

[[0.71088435 0.2244898 0.06462585]

[0.12693012 0.84637529 0.02669458]

[0.06294537 0.05581948 0.88123515]]





precision recall f1-score support

hate 0.28 0.71 0.40 294

offensive 0.97 0.85 0.90 3821

neither 0.86 0.88 0.87 842

accuracy 0.84 4957

macro avg 0.70 0.81 0.72 4957

weighted avg 0.91 0.84 0.87 4957

average eval\_loss: 0.8120663950572752, accuracy: 0.8442606415170466