hate as offensive

alpha=0.05

1 run...

Downloading: 100%

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

Downloading: 100%

263M/263M [00:08<00:00, 32.7MB/s]

training 1 epoch...

45.300206661224365 seconds used......

1 training finished! average train loss: 0.7029988781401986

total clf loss: 123.29289514571428 total adv loss: 61.601748052984476

evaluating...

Confusion matrix, without normalization

[[ 186 83 25]

[ 331 3358 132]

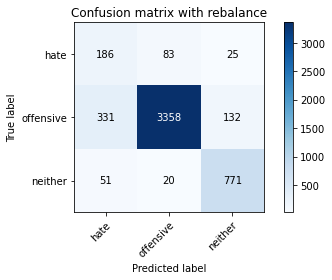
[ 51 20 771]]

Normalized confusion matrix

[[0.63265306 0.28231293 0.08503401]

[0.08662654 0.87882753 0.03454593]

[0.06057007 0.02375297 0.91567696]]





precision recall f1-score support

hate 0.33 0.63 0.43 294

offensive 0.97 0.88 0.92 3821

neither 0.83 0.92 0.87 842

accuracy 0.87 4957

macro avg 0.71 0.81 0.74 4957

weighted avg 0.91 0.87 0.88 4957

average eval\_loss: 0.38881737744856265, accuracy: 0.8704861811579584

training 2 epoch...

44.74645256996155 seconds used......

2 training finished! average train loss: 0.40008783039816637

total clf loss: 69.86523019522429 total adv loss: 29.004226684570312

evaluating...

Confusion matrix, without normalization

[[ 203 81 10]

[ 374 3372 75]

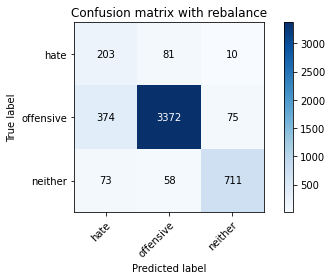
[ 73 58 711]]

Normalized confusion matrix

[[0.69047619 0.2755102 0.03401361]

[0.09788014 0.88249149 0.01962837]

[0.08669834 0.06888361 0.84441805]]





precision recall f1-score support

hate 0.31 0.69 0.43 294

offensive 0.96 0.88 0.92 3821

neither 0.89 0.84 0.87 842

accuracy 0.86 4957

macro avg 0.72 0.81 0.74 4957

weighted avg 0.91 0.86 0.88 4957

average eval\_loss: 0.36036917119617423, accuracy: 0.864635868468832

training 3 epoch...

44.83049535751343 seconds used......

3 training finished! average train loss: 0.27504383238871194

total clf loss: 47.857174437493086 total adv loss: 16.493583233095706

evaluating...

Confusion matrix, without normalization

[[ 238 40 16]

[ 953 2768 100]

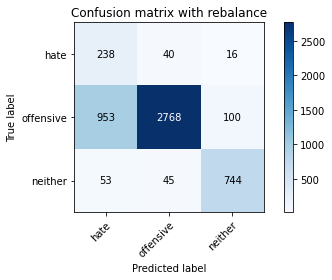
[ 53 45 744]]

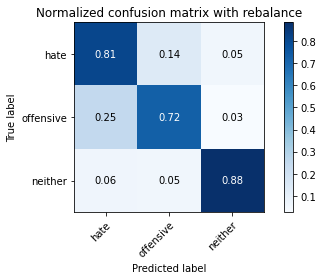
Normalized confusion matrix

[[0.80952381 0.13605442 0.05442177]

[0.24941115 0.72441769 0.02617116]

[0.06294537 0.05344418 0.88361045]]





precision recall f1-score support

hate 0.19 0.81 0.31 294

offensive 0.97 0.72 0.83 3821

neither 0.87 0.88 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: 0.5629300568853656, accuracy: 0.7565059511801493

training 4 epoch...

44.80996584892273 seconds used......

4 training finished! average train loss: 0.15814788578546535

total clf loss: 27.347926313057542 total adv loss: 6.092756372701842

evaluating...

Confusion matrix, without normalization

[[ 239 42 13]

[ 977 2748 96]

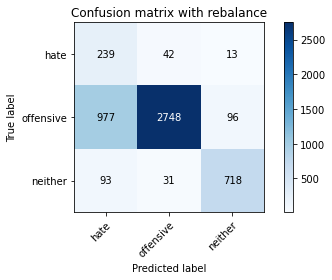
[ 93 31 718]]

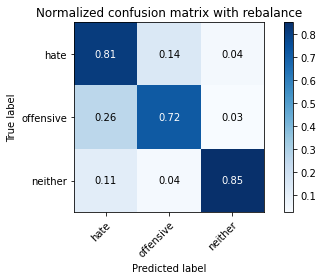
Normalized confusion matrix

[[0.81292517 0.14285714 0.04421769]

[0.25569223 0.71918346 0.02512431]

[0.11045131 0.0368171 0.85273159]]





precision recall f1-score support

hate 0.18 0.81 0.30 294

offensive 0.97 0.72 0.83 3821

neither 0.87 0.85 0.86 842

accuracy 0.75 4957

macro avg 0.67 0.79 0.66 4957

weighted avg 0.91 0.75 0.80 4957

average eval\_loss: 0.6711201781046486, accuracy: 0.7474278797659875

training 5 epoch...

44.800111293792725 seconds used......

5 training finished! average train loss: 0.09111201698886372

total clf loss: 15.704313167370856 total adv loss: 2.4831656282949552

evaluating...

Confusion matrix, without normalization

[[ 230 49 15]

[ 900 2808 113]

[ 69 27 746]]

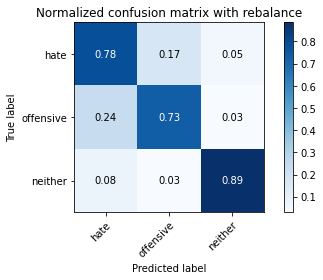
Normalized confusion matrix

[[0.78231293 0.16666667 0.05102041]

[0.23554043 0.73488616 0.02957341]

[0.08194774 0.03206651 0.88598575]]





precision recall f1-score support

hate 0.19 0.78 0.31 294

offensive 0.97 0.73 0.84 3821

neither 0.85 0.89 0.87 842

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: 0.8168097625696852, accuracy: 0.7633649384708493

training 6 epoch...

44.80152368545532 seconds used......

6 training finished! average train loss: 0.050488557140628824

total clf loss: 8.75831608613953 total adv loss: 2.4954559096986486

evaluating...

Confusion matrix, without normalization

[[ 223 57 14]

[ 707 2959 155]

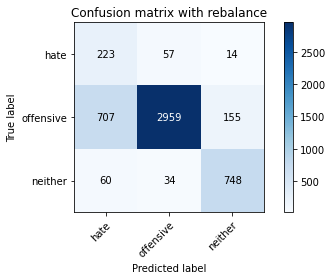
[ 60 34 748]]

Normalized confusion matrix

[[0.7585034 0.19387755 0.04761905]

[0.1850301 0.77440461 0.0405653 ]

[0.07125891 0.04038005 0.88836105]]





precision recall f1-score support

hate 0.23 0.76 0.35 294

offensive 0.97 0.77 0.86 3821

neither 0.82 0.89 0.85 842

accuracy 0.79 4957

macro avg 0.67 0.81 0.69 4957

weighted avg 0.90 0.79 0.83 4957

average eval\_loss: 0.8405151234398927, accuracy: 0.7928182368367964

training 7 epoch...

44.747451066970825 seconds used......

7 training finished! average train loss: 0.022096949915417975

total clf loss: 3.828630296047777 total adv loss: 1.0010372598403592

evaluating...

Confusion matrix, without normalization

[[ 210 61 23]

[ 611 3039 171]

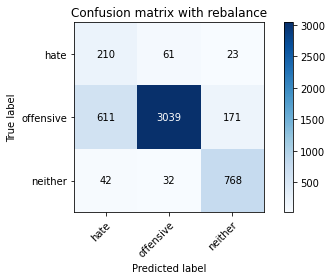
[ 42 32 768]]

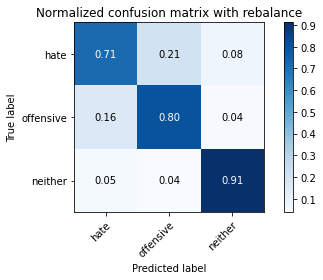
Normalized confusion matrix

[[0.71428571 0.20748299 0.07823129]

[0.15990578 0.79534153 0.04475268]

[0.04988124 0.03800475 0.91211401]]





precision recall f1-score support

hate 0.24 0.71 0.36 294

offensive 0.97 0.80 0.87 3821

neither 0.80 0.91 0.85 842

accuracy 0.81 4957

macro avg 0.67 0.81 0.70 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.9284004467809873, accuracy: 0.8103691749041759

training 8 epoch...

44.75951623916626 seconds used......

8 training finished! average train loss: 0.02976332520398894

total clf loss: 5.298203610116616 total adv loss: 4.173500341938961

evaluating...

Confusion matrix, without normalization

[[ 199 77 18]

[ 589 3115 117]

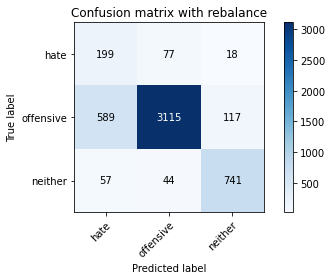
[ 57 44 741]]

Normalized confusion matrix

[[0.67687075 0.26190476 0.06122449]

[0.15414813 0.81523161 0.03062026]

[0.06769596 0.05225653 0.88004751]]





precision recall f1-score support

hate 0.24 0.68 0.35 294

offensive 0.96 0.82 0.88 3821

neither 0.85 0.88 0.86 842

accuracy 0.82 4957

macro avg 0.68 0.79 0.70 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.8140867793662173, accuracy: 0.8180351018761347

training 9 epoch...

44.71815252304077 seconds used......

9 training finished! average train loss: 0.01245752208503998

total clf loss: 2.2180645680928137 total adv loss: 1.7565655494690873

evaluating...

Confusion matrix, without normalization

[[ 224 55 15]

[ 893 2802 126]

[ 74 37 731]]

Normalized confusion matrix

[[0.76190476 0.18707483 0.05102041]

[0.23370845 0.73331589 0.03297566]

[0.08788599 0.04394299 0.86817102]]





precision recall f1-score support

hate 0.19 0.76 0.30 294

offensive 0.97 0.73 0.83 3821

neither 0.84 0.87 0.85 842

accuracy 0.76 4957

macro avg 0.66 0.79 0.66 4957

weighted avg 0.90 0.76 0.81 4957

average eval\_loss: 1.2296341335731409, accuracy: 0.7579180956223522

training 10 epoch...

44.787365436553955 seconds used......

10 training finished! average train loss: 0.03210926437644936

total clf loss: 5.61058009948465 total adv loss: 2.397918042486708

evaluating...

Confusion matrix, without normalization

[[ 202 72 20]

[ 585 3125 111]

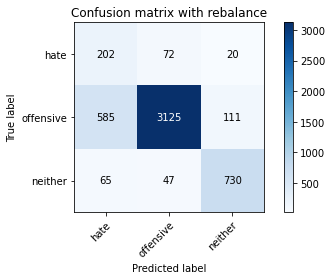
[ 65 47 730]]

Normalized confusion matrix

[[0.68707483 0.24489796 0.06802721]

[0.15310128 0.81784873 0.02904999]

[0.07719715 0.05581948 0.86698337]]





precision recall f1-score support

hate 0.24 0.69 0.35 294

offensive 0.96 0.82 0.88 3821

neither 0.85 0.87 0.86 842

accuracy 0.82 4957

macro avg 0.68 0.79 0.70 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.8370932676311161, accuracy: 0.8184385717167642

2 run...

training 1 epoch...

44.82785606384277 seconds used......

1 training finished! average train loss: 0.6457180909768879

total clf loss: 113.19224609434605 total adv loss: 55.48905062675476

evaluating...

Confusion matrix, without normalization

[[ 234 41 19]

[ 826 2898 97]

[ 98 21 723]]

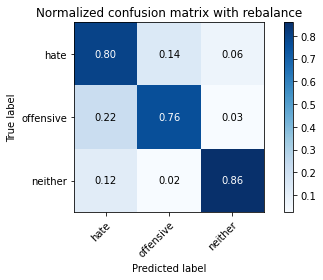
Normalized confusion matrix

[[0.79591837 0.13945578 0.06462585]

[0.21617378 0.7584402 0.02538602]

[0.11638955 0.02494062 0.85866983]]





precision recall f1-score support

hate 0.20 0.80 0.32 294

offensive 0.98 0.76 0.85 3821

neither 0.86 0.86 0.86 842

accuracy 0.78 4957

macro avg 0.68 0.80 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.5020619768408998, accuracy: 0.7776881178131935

training 2 epoch...

44.970399141311646 seconds used......

2 training finished! average train loss: 0.40418653255491926

total clf loss: 70.62292122840881 total adv loss: 30.14048113953322

evaluating...

Confusion matrix, without normalization

[[ 193 69 32]

[ 469 3198 154]

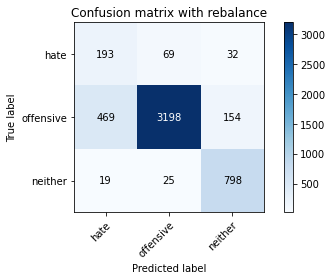
[ 19 25 798]]

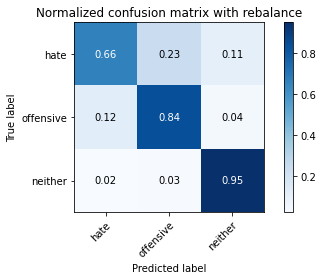
Normalized confusion matrix

[[0.65646259 0.23469388 0.10884354]

[0.12274274 0.83695368 0.04030359]

[0.02256532 0.02969121 0.94774347]]





precision recall f1-score support

hate 0.28 0.66 0.40 294

offensive 0.97 0.84 0.90 3821

neither 0.81 0.95 0.87 842

accuracy 0.85 4957

macro avg 0.69 0.81 0.72 4957

weighted avg 0.90 0.85 0.87 4957

average eval\_loss: 0.3827595985945194, accuracy: 0.8450675811983054

training 3 epoch...

44.99462938308716 seconds used......

3 training finished! average train loss: 0.2804517406120635

total clf loss: 48.74810640886426 total adv loss: 15.817176394863054

evaluating...

Confusion matrix, without normalization

[[ 229 45 20]

[ 729 2957 135]

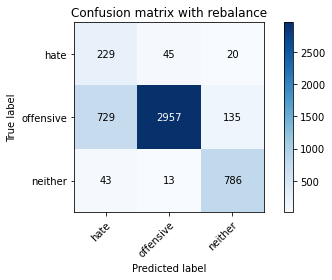
[ 43 13 786]]

Normalized confusion matrix

[[0.77891156 0.15306122 0.06802721]

[0.19078775 0.77388118 0.03533107]

[0.05106888 0.01543943 0.93349169]]





precision recall f1-score support

hate 0.23 0.78 0.35 294

offensive 0.98 0.77 0.87 3821

neither 0.84 0.93 0.88 842

accuracy 0.80 4957

macro avg 0.68 0.83 0.70 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.5449373944692554, accuracy: 0.8012911034900141

training 4 epoch...

44.786309242248535 seconds used......

4 training finished! average train loss: 0.1571148359897541

total clf loss: 27.153387188911438 total adv loss: 5.735003475681879

evaluating...

Confusion matrix, without normalization

[[ 226 52 16]

[ 917 2786 118]

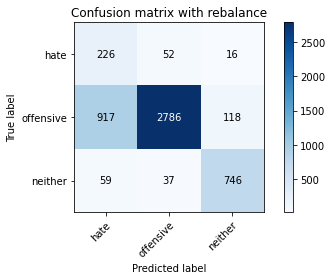
[ 59 37 746]]

Normalized confusion matrix

[[0.76870748 0.17687075 0.05442177]

[0.23998953 0.7291285 0.03088197]

[0.07007126 0.04394299 0.88598575]]





precision recall f1-score support

hate 0.19 0.77 0.30 294

offensive 0.97 0.73 0.83 3821

neither 0.85 0.89 0.87 842

accuracy 0.76 4957

macro avg 0.67 0.79 0.67 4957

weighted avg 0.90 0.76 0.81 4957

average eval\_loss: 0.6400259241101242, accuracy: 0.758119830542667

training 5 epoch...

44.97901010513306 seconds used......

5 training finished! average train loss: 0.07784809917756165

total clf loss: 13.48711299430579 total adv loss: 3.5017603428132134

evaluating...

Confusion matrix, without normalization

[[ 227 54 13]

[ 653 3075 93]

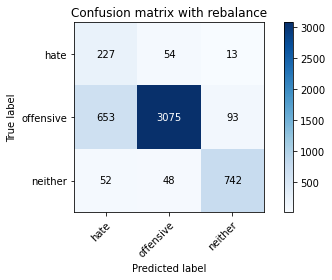
[ 52 48 742]]

Normalized confusion matrix

[[0.77210884 0.18367347 0.04421769]

[0.17089767 0.80476315 0.02433918]

[0.06175772 0.05700713 0.88123515]]





precision recall f1-score support

hate 0.24 0.77 0.37 294

offensive 0.97 0.80 0.88 3821

neither 0.88 0.88 0.88 842

accuracy 0.82 4957

macro avg 0.70 0.82 0.71 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.6671876091464993, accuracy: 0.815816017752673

training 6 epoch...

44.964778661727905 seconds used......

6 training finished! average train loss: 0.03795195834395307

total clf loss: 6.565883020521142 total adv loss: 1.5219636382507815

evaluating...

Confusion matrix, without normalization

[[ 225 46 23]

[ 885 2804 132]

[ 42 29 771]]

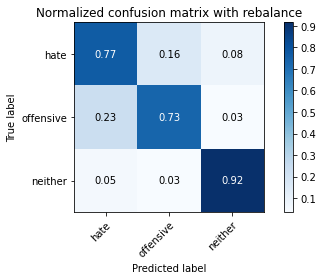
Normalized confusion matrix

[[0.76530612 0.15646259 0.07823129]

[0.23161476 0.73383931 0.03454593]

[0.04988124 0.03444181 0.91567696]]





precision recall f1-score support

hate 0.20 0.77 0.31 294

offensive 0.97 0.73 0.84 3821

neither 0.83 0.92 0.87 842

accuracy 0.77 4957

macro avg 0.67 0.80 0.67 4957

weighted avg 0.90 0.77 0.81 4957

average eval\_loss: 0.9985949210522156, accuracy: 0.7665926971958846

training 7 epoch...

44.88895106315613 seconds used......

7 training finished! average train loss: 0.030134432421172072

total clf loss: 5.229426359874196 total adv loss: 1.5287676941134123

evaluating...

Confusion matrix, without normalization

[[ 242 41 11]

[1051 2689 81]

[ 124 49 669]]

Normalized confusion matrix

[[0.82312925 0.13945578 0.03741497]

[0.27505889 0.70374248 0.02119864]

[0.14726841 0.05819477 0.79453682]]





precision recall f1-score support

hate 0.17 0.82 0.28 294

offensive 0.97 0.70 0.81 3821

neither 0.88 0.79 0.83 842

accuracy 0.73 4957

macro avg 0.67 0.77 0.64 4957

weighted avg 0.91 0.73 0.79 4957

average eval\_loss: 1.2746719380299891, accuracy: 0.7262457131329433

training 8 epoch...

44.84750151634216 seconds used......

8 training finished! average train loss: 0.014423524575887999

total clf loss: 2.4855017802328803 total adv loss: 0.3815812094803732

evaluating...

Confusion matrix, without normalization

[[ 221 59 14]

[ 720 3012 89]

[ 59 68 715]]

Normalized confusion matrix

[[0.75170068 0.20068027 0.04761905]

[0.18843235 0.78827532 0.02329233]

[0.07007126 0.0807601 0.84916865]]





precision recall f1-score support

hate 0.22 0.75 0.34 294

offensive 0.96 0.79 0.87 3821

neither 0.87 0.85 0.86 842

accuracy 0.80 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.90 0.80 0.83 4957

average eval\_loss: 0.9737714456277899, accuracy: 0.7964494654024612

training 9 epoch...

44.847697257995605 seconds used......

9 training finished! average train loss: 0.020144512465414974

total clf loss: 3.45609786894056 total adv loss: 0.22772488801061286

evaluating...

Confusion matrix, without normalization

[[ 231 48 15]

[ 695 3021 105]

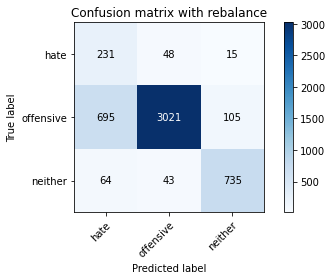
[ 64 43 735]]

Normalized confusion matrix

[[0.78571429 0.16326531 0.05102041]

[0.18188956 0.79063072 0.02747972]

[0.0760095 0.05106888 0.87292162]]





precision recall f1-score support

hate 0.23 0.79 0.36 294

offensive 0.97 0.79 0.87 3821

neither 0.86 0.87 0.87 842

accuracy 0.80 4957

macro avg 0.69 0.82 0.70 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.9457989693159282, accuracy: 0.8043171272947347

training 10 epoch...

45.026899099349976 seconds used......

10 training finished! average train loss: 0.010004228356087374

total clf loss: 1.7656181880738586 total adv loss: 1.097902872220999

evaluating...

Confusion matrix, without normalization

[[ 215 60 19]

[ 671 3054 96]

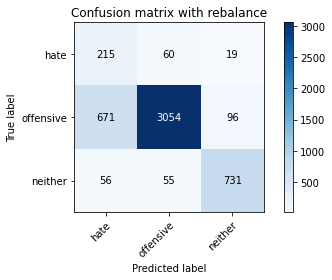
[ 56 55 731]]

Normalized confusion matrix

[[0.73129252 0.20408163 0.06462585]

[0.17560848 0.79926721 0.02512431]

[0.06650831 0.06532067 0.86817102]]





precision recall f1-score support

hate 0.23 0.73 0.35 294

offensive 0.96 0.80 0.87 3821

neither 0.86 0.87 0.87 842

accuracy 0.81 4957

macro avg 0.69 0.80 0.70 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.9842152623572357, accuracy: 0.8069396812588259

3 run...

training 1 epoch...

44.961037397384644 seconds used......

1 training finished! average train loss: 0.6789071674932513

total clf loss: 119.15427435934544 total adv loss: 61.222975039854646

evaluating...

Confusion matrix, without normalization

[[ 221 42 31]

[ 746 2900 175]

[ 24 18 800]]

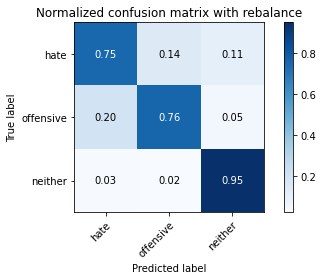
Normalized confusion matrix

[[0.75170068 0.14285714 0.10544218]

[0.19523685 0.75896362 0.04579953]

[0.02850356 0.02137767 0.95011876]]





precision recall f1-score support

hate 0.22 0.75 0.34 294

offensive 0.98 0.76 0.86 3821

neither 0.80 0.95 0.87 842

accuracy 0.79 4957

macro avg 0.67 0.82 0.69 4957

weighted avg 0.90 0.79 0.83 4957

average eval\_loss: 0.4969225738558077, accuracy: 0.7910026225539641

training 2 epoch...

44.75889587402344 seconds used......

2 training finished! average train loss: 0.3894097773216621

total clf loss: 67.95558244735003 total adv loss: 27.330205141566694

evaluating...

Confusion matrix, without normalization

[[ 237 39 18]

[ 800 2891 130]

[ 38 20 784]]

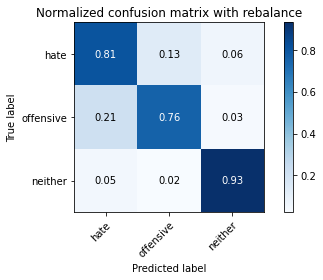
Normalized confusion matrix

[[0.80612245 0.13265306 0.06122449]

[0.20936928 0.75660822 0.03402251]

[0.04513064 0.02375297 0.93111639]]





precision recall f1-score support

hate 0.22 0.81 0.35 294

offensive 0.98 0.76 0.85 3821

neither 0.84 0.93 0.88 842

accuracy 0.79 4957

macro avg 0.68 0.83 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.4935380387450418, accuracy: 0.7891870082711318

training 3 epoch...

44.765342235565186 seconds used......

3 training finished! average train loss: 0.2672453976906182

total clf loss: 46.34065004438162 total adv loss: 12.833743941679131

evaluating...

Confusion matrix, without normalization

[[ 231 45 18]

[ 805 2888 128]

[ 48 22 772]]

Normalized confusion matrix

[[0.78571429 0.15306122 0.06122449]

[0.21067783 0.75582308 0.03349908]

[0.05700713 0.02612827 0.91686461]]





precision recall f1-score support

hate 0.21 0.79 0.34 294

offensive 0.98 0.76 0.85 3821

neither 0.84 0.92 0.88 842

accuracy 0.78 4957

macro avg 0.68 0.82 0.69 4957

weighted avg 0.91 0.78 0.83 4957

average eval\_loss: 0.5501055075456539, accuracy: 0.7849505749445229

training 4 epoch...

44.791264057159424 seconds used......

4 training finished! average train loss: 0.15830386038433913

total clf loss: 27.30788946337998 total adv loss: 4.758584791743488

evaluating...

Confusion matrix, without normalization

[[ 244 40 10]

[1132 2572 117]

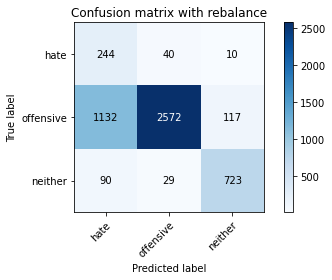
[ 90 29 723]]

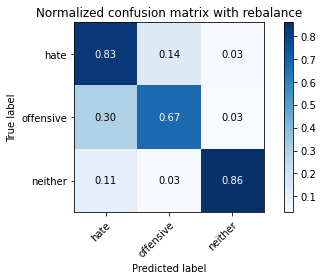
Normalized confusion matrix

[[0.82993197 0.13605442 0.03401361]

[0.29625752 0.67312222 0.03062026]

[0.10688836 0.03444181 0.85866983]]





precision recall f1-score support

hate 0.17 0.83 0.28 294

offensive 0.97 0.67 0.80 3821

neither 0.85 0.86 0.85 842

accuracy 0.71 4957

macro avg 0.66 0.79 0.64 4957

weighted avg 0.91 0.71 0.78 4957

average eval\_loss: 0.7950657305397815, accuracy: 0.7139398829937462

training 5 epoch...

44.90039038658142 seconds used......

5 training finished! average train loss: 0.08074419444416002

total clf loss: 13.951756528578699 total adv loss: 2.8899853851862645

evaluating...

Confusion matrix, without normalization

[[ 208 67 19]

[ 518 3162 141]

[ 44 34 764]]

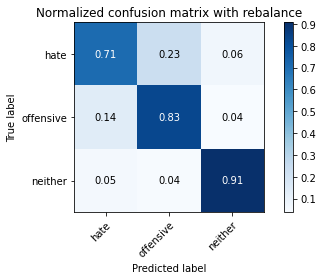
Normalized confusion matrix

[[0.70748299 0.22789116 0.06462585]

[0.13556661 0.82753206 0.03690133]

[0.05225653 0.04038005 0.90736342]]





precision recall f1-score support

hate 0.27 0.71 0.39 294

offensive 0.97 0.83 0.89 3821

neither 0.83 0.91 0.87 842

accuracy 0.83 4957

macro avg 0.69 0.81 0.72 4957

weighted avg 0.90 0.83 0.86 4957

average eval\_loss: 0.6128605763472977, accuracy: 0.8339721605809965

training 6 epoch...

44.862284660339355 seconds used......

6 training finished! average train loss: 0.048751966808568574

total clf loss: 8.420436853310093 total adv loss: 1.6770111763236173

evaluating...

Confusion matrix, without normalization

[[ 198 69 27]

[ 557 3123 141]

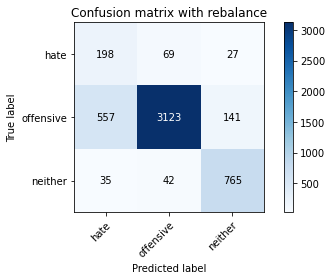
[ 35 42 765]]

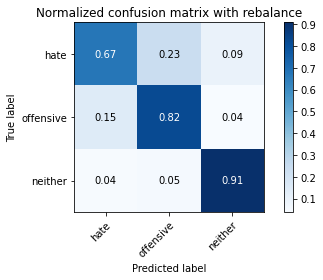
Normalized confusion matrix

[[0.67346939 0.23469388 0.09183673]

[0.14577336 0.81732531 0.03690133]

[0.0415677 0.04988124 0.90855107]]





precision recall f1-score support

hate 0.25 0.67 0.37 294

offensive 0.97 0.82 0.89 3821

neither 0.82 0.91 0.86 842

accuracy 0.82 4957

macro avg 0.68 0.80 0.70 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.7353835641678738, accuracy: 0.8242888844058907

training 7 epoch...

44.887837409973145 seconds used......

7 training finished! average train loss: 0.0243120703530329

total clf loss: 4.212673387723044 total adv loss: 1.1061873102339632

evaluating...

Confusion matrix, without normalization

[[ 225 51 18]

[ 832 2877 112]

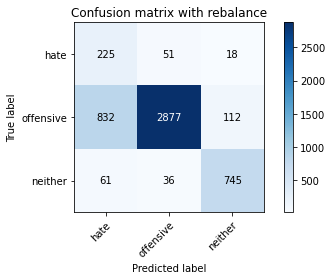
[ 61 36 745]]

Normalized confusion matrix

[[0.76530612 0.17346939 0.06122449]

[0.21774405 0.75294426 0.0293117 ]

[0.07244656 0.04275534 0.8847981 ]]





precision recall f1-score support

hate 0.20 0.77 0.32 294

offensive 0.97 0.75 0.85 3821

neither 0.85 0.88 0.87 842

accuracy 0.78 4957

macro avg 0.67 0.80 0.68 4957

weighted avg 0.90 0.78 0.82 4957

average eval\_loss: 1.0772805486114756, accuracy: 0.7760742384506758

training 8 epoch...

45.04916071891785 seconds used......

8 training finished! average train loss: 0.023849978994160116

total clf loss: 4.136023912840756 total adv loss: 1.1535497288068086

evaluating...

Confusion matrix, without normalization

[[ 236 43 15]

[ 883 2819 119]

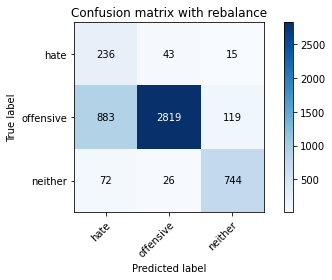
[ 72 26 744]]

Normalized confusion matrix

[[0.80272109 0.1462585 0.05102041]

[0.23109134 0.73776498 0.03114368]

[0.08551069 0.03087886 0.88361045]]





precision recall f1-score support

hate 0.20 0.80 0.32 294

offensive 0.98 0.74 0.84 3821

neither 0.85 0.88 0.87 842

accuracy 0.77 4957

macro avg 0.67 0.81 0.67 4957

weighted avg 0.91 0.77 0.81 4957

average eval\_loss: 1.2689976057698649, accuracy: 0.7663909622755699

training 9 epoch...

44.80662798881531 seconds used......

9 training finished! average train loss: 0.01609195743031168

total clf loss: 2.7826617992541287 total adv loss: 0.6187414039066539

evaluating...

Confusion matrix, without normalization

[[ 218 59 17]

[ 600 3121 100]

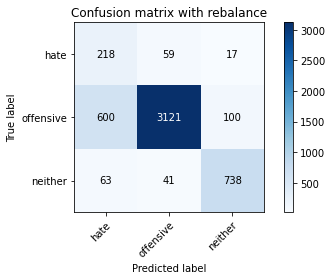
[ 63 41 738]]

Normalized confusion matrix

[[0.7414966 0.20068027 0.05782313]

[0.15702696 0.81680188 0.02617116]

[0.07482185 0.04869359 0.87648456]]





precision recall f1-score support

hate 0.25 0.74 0.37 294

offensive 0.97 0.82 0.89 3821

neither 0.86 0.88 0.87 842

accuracy 0.82 4957

macro avg 0.69 0.81 0.71 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.9148927665617873, accuracy: 0.8224732701230583

training 10 epoch...

44.851826429367065 seconds used......

10 training finished! average train loss: 0.015395571223180332

total clf loss: 2.67405328861787 total adv loss: 0.8282121424158504

evaluating...

Confusion matrix, without normalization

[[ 217 59 18]

[ 612 3106 103]

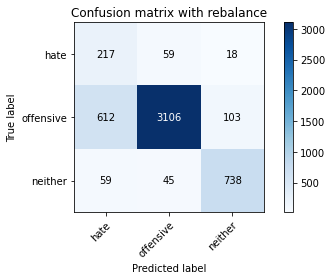
[ 59 45 738]]

Normalized confusion matrix

[[0.73809524 0.20068027 0.06122449]

[0.1601675 0.81287621 0.02695629]

[0.07007126 0.05344418 0.87648456]]





precision recall f1-score support

hate 0.24 0.74 0.37 294

offensive 0.97 0.81 0.88 3821

neither 0.86 0.88 0.87 842

accuracy 0.82 4957

macro avg 0.69 0.81 0.71 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.9233790453731818, accuracy: 0.819245511398023

4 run...

training 1 epoch...

44.81007981300354 seconds used......

1 training finished! average train loss: 0.648556101653311

total clf loss: 113.70166848599911 total adv loss: 55.97149858996272

evaluating...

Confusion matrix, without normalization

[[ 244 31 19]

[1002 2708 111]

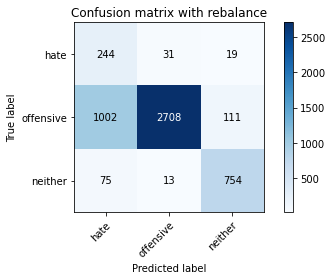
[ 75 13 754]]

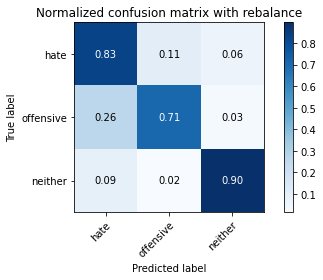
Normalized confusion matrix

[[0.82993197 0.10544218 0.06462585]

[0.26223502 0.708715 0.02904999]

[0.08907363 0.01543943 0.89548694]]





precision recall f1-score support

hate 0.18 0.83 0.30 294

offensive 0.98 0.71 0.82 3821

neither 0.85 0.90 0.87 842

accuracy 0.75 4957

macro avg 0.67 0.81 0.67 4957

weighted avg 0.91 0.75 0.80 4957

average eval\_loss: 0.621316623303198, accuracy: 0.7476296146863022

training 2 epoch...

44.802809715270996 seconds used......

2 training finished! average train loss: 0.4078700688411618

total clf loss: 71.15231415629387 total adv loss: 28.130650366656482

evaluating...

Confusion matrix, without normalization

[[ 216 42 36]

[ 700 2967 154]

[ 16 24 802]]

Normalized confusion matrix

[[0.73469388 0.14285714 0.12244898]

[0.18319812 0.7764983 0.04030359]

[0.01900238 0.02850356 0.95249406]]





precision recall f1-score support

hate 0.23 0.73 0.35 294

offensive 0.98 0.78 0.87 3821

neither 0.81 0.95 0.87 842

accuracy 0.80 4957

macro avg 0.67 0.82 0.70 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.4630608957921786, accuracy: 0.8039136574541053

training 3 epoch...

44.81328201293945 seconds used......

3 training finished! average train loss: 0.28168417173519467

total clf loss: 49.00518328696489 total adv loss: 16.74379561585374

evaluating...

Confusion matrix, without normalization

[[ 234 41 19]

[ 765 2941 115]

[ 46 33 763]]

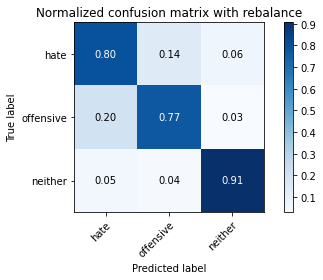
Normalized confusion matrix

[[0.79591837 0.13945578 0.06462585]

[0.20020937 0.7696938 0.03009683]

[0.05463183 0.0391924 0.90617577]]





precision recall f1-score support

hate 0.22 0.80 0.35 294

offensive 0.98 0.77 0.86 3821

neither 0.85 0.91 0.88 842

accuracy 0.79 4957

macro avg 0.68 0.82 0.70 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.5025558454705582, accuracy: 0.7944321161993141

training 4 epoch...

44.875078439712524 seconds used......

4 training finished! average train loss: 0.1683041154357948

total clf loss: 29.140186801552773 total adv loss: 7.203661916631972

evaluating...

Confusion matrix, without normalization

[[ 238 43 13]

[ 826 2890 105]

[ 55 38 749]]

Normalized confusion matrix

[[0.80952381 0.1462585 0.04421769]

[0.21617378 0.75634651 0.02747972]

[0.06532067 0.04513064 0.88954869]]





precision recall f1-score support

hate 0.21 0.81 0.34 294

offensive 0.97 0.76 0.85 3821

neither 0.86 0.89 0.88 842

accuracy 0.78 4957

macro avg 0.68 0.82 0.69 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.5967492726331036, accuracy: 0.782126286060117

training 5 epoch...

44.96721005439758 seconds used......

5 training finished! average train loss: 0.082619670869739

total clf loss: 14.22845147899352 total adv loss: 2.0097550195387157

evaluating...

Confusion matrix, without normalization

[[ 211 66 17]

[ 645 3073 103]

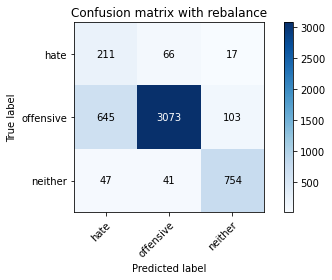
[ 47 41 754]]

Normalized confusion matrix

[[0.71768707 0.2244898 0.05782313]

[0.16880398 0.80423973 0.02695629]

[0.05581948 0.04869359 0.89548694]]





precision recall f1-score support

hate 0.23 0.72 0.35 294

offensive 0.97 0.80 0.88 3821

neither 0.86 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: 0.6313615631004195, accuracy: 0.8146056082307848

training 6 epoch...

44.85893130302429 seconds used......

6 training finished! average train loss: 0.04781516521014491

total clf loss: 8.218617215519771 total adv loss: 0.8444795038058146

evaluating...

Confusion matrix, without normalization

[[ 219 62 13]

[ 595 3114 112]

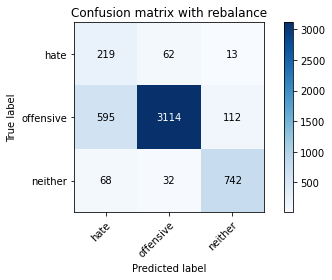
[ 68 32 742]]

Normalized confusion matrix

[[0.74489796 0.21088435 0.04421769]

[0.1557184 0.8149699 0.0293117 ]

[0.0807601 0.03800475 0.88123515]]





precision recall f1-score support

hate 0.25 0.74 0.37 294

offensive 0.97 0.81 0.89 3821

neither 0.86 0.88 0.87 842

accuracy 0.82 4957

macro avg 0.69 0.81 0.71 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.6951240973824996, accuracy: 0.8220698002824289

training 7 epoch...

44.77983283996582 seconds used......

7 training finished! average train loss: 0.041999671883249316

total clf loss: 7.378321189782582 total adv loss: 3.9275461926890785

evaluating...

Confusion matrix, without normalization

[[ 212 63 19]

[ 723 2979 119]

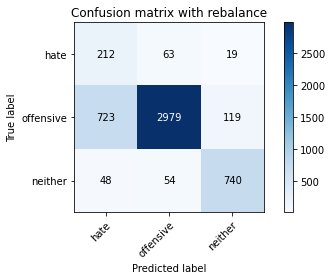
[ 48 54 740]]

Normalized confusion matrix

[[0.72108844 0.21428571 0.06462585]

[0.18921748 0.77963884 0.03114368]

[0.05700713 0.06413302 0.87885986]]





precision recall f1-score support

hate 0.22 0.72 0.33 294

offensive 0.96 0.78 0.86 3821

neither 0.84 0.88 0.86 842

accuracy 0.79 4957

macro avg 0.67 0.79 0.68 4957

weighted avg 0.90 0.79 0.83 4957

average eval\_loss: 0.8647174894914872, accuracy: 0.7930199717571111

training 8 epoch...

44.83740520477295 seconds used......

8 training finished! average train loss: 0.02221594977925112

total clf loss: 3.858876567246625 total adv loss: 1.1989830863035422

evaluating...

Confusion matrix, without normalization

[[ 220 57 17]

[ 658 3022 141]

[ 44 28 770]]

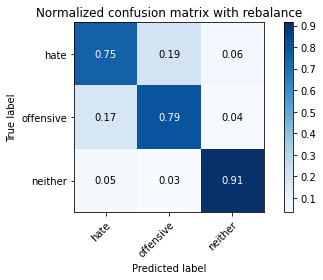
Normalized confusion matrix

[[0.74829932 0.19387755 0.05782313]

[0.17220623 0.79089244 0.03690133]

[0.05225653 0.03325416 0.91448931]]





precision recall f1-score support

hate 0.24 0.75 0.36 294

offensive 0.97 0.79 0.87 3821

neither 0.83 0.91 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.82 0.70 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.9571205815769762, accuracy: 0.8093605003026024

training 9 epoch...

44.7162652015686 seconds used......

9 training finished! average train loss: 0.005872821304365061

total clf loss: 1.023536578839412 total adv loss: 0.3856828415828204

evaluating...

Confusion matrix, without normalization

[[ 224 55 15]

[ 641 3078 102]

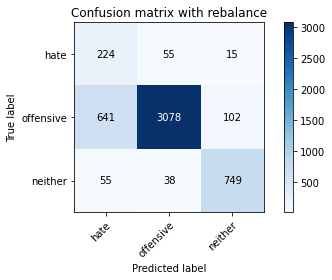
[ 55 38 749]]

Normalized confusion matrix

[[0.76190476 0.18707483 0.05102041]

[0.16775713 0.80554829 0.02669458]

[0.06532067 0.04513064 0.88954869]]





precision recall f1-score support

hate 0.24 0.76 0.37 294

offensive 0.97 0.81 0.88 3821

neither 0.86 0.89 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: 1.0359274644006584, accuracy: 0.8172281621948759

training 10 epoch...

44.77485728263855 seconds used......

10 training finished! average train loss: 0.016285592220789666

total clf loss: 2.865504965215223 total adv loss: 1.6133738564756186

evaluating...

Confusion matrix, without normalization

[[ 224 54 16]

[ 734 2994 93]

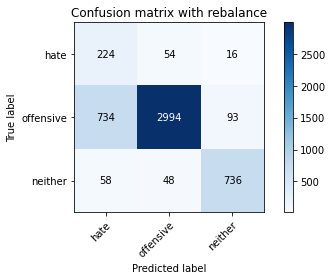
[ 58 48 736]]

Normalized confusion matrix

[[0.76190476 0.18367347 0.05442177]

[0.19209631 0.78356451 0.02433918]

[0.06888361 0.05700713 0.87410926]]





precision recall f1-score support

hate 0.22 0.76 0.34 294

offensive 0.97 0.78 0.87 3821

neither 0.87 0.87 0.87 842

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: 1.068856152949945, accuracy: 0.7976598749243494

5 run...

training 1 epoch...

44.81622314453125 seconds used......

1 training finished! average train loss: 0.6625412098671261

total clf loss: 116.3138466924429 total adv loss: 60.385998986661434

evaluating...

Confusion matrix, without normalization

[[ 228 45 21]

[ 861 2862 98]

[ 67 19 756]]

Normalized confusion matrix

[[0.7755102 0.15306122 0.07142857]

[0.22533368 0.74901858 0.02564774]

[0.07957245 0.02256532 0.89786223]]





precision recall f1-score support

hate 0.20 0.78 0.31 294

offensive 0.98 0.75 0.85 3821

neither 0.86 0.90 0.88 842

accuracy 0.78 4957

macro avg 0.68 0.81 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.5300125325278889, accuracy: 0.7758725035303611

training 2 epoch...

44.780837297439575 seconds used......

2 training finished! average train loss: 0.3979523212414736

total clf loss: 69.39361450821161 total adv loss: 26.87535112723708

evaluating...

Confusion matrix, without normalization

[[ 236 46 12]

[ 797 2945 79]

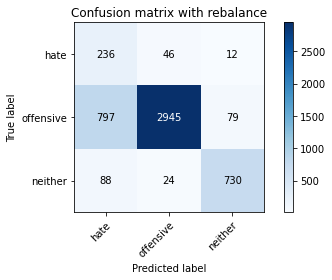
[ 88 24 730]]

Normalized confusion matrix

[[0.80272109 0.15646259 0.04081633]

[0.20858414 0.77074064 0.02067522]

[0.10451306 0.02850356 0.86698337]]





precision recall f1-score support

hate 0.21 0.80 0.33 294

offensive 0.98 0.77 0.86 3821

neither 0.89 0.87 0.88 842

accuracy 0.79 4957

macro avg 0.69 0.81 0.69 4957

weighted avg 0.92 0.79 0.83 4957

average eval\_loss: 0.5025660110817801, accuracy: 0.788985273350817

training 3 epoch...

44.777093172073364 seconds used......

3 training finished! average train loss: 0.27377078825967355

total clf loss: 47.49377777427435 total adv loss: 13.579459807719104

evaluating...

Confusion matrix, without normalization

[[ 216 59 19]

[ 754 2957 110]

[ 51 27 764]]

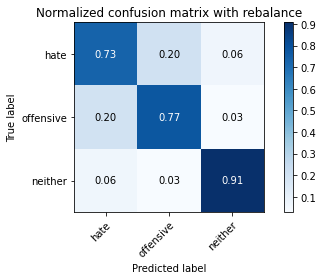
Normalized confusion matrix

[[0.73469388 0.20068027 0.06462585]

[0.19733054 0.77388118 0.02878828]

[0.06057007 0.03206651 0.90736342]]





precision recall f1-score support

hate 0.21 0.73 0.33 294

offensive 0.97 0.77 0.86 3821

neither 0.86 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: 0.4796604356998878, accuracy: 0.7942303812789994

training 4 epoch...

44.7406222820282 seconds used......

4 training finished! average train loss: 0.16698953502795152

total clf loss: 28.83776799403131 total adv loss: 5.651148303761147

evaluating...

Confusion matrix, without normalization

[[ 191 82 21]

[ 466 3254 101]

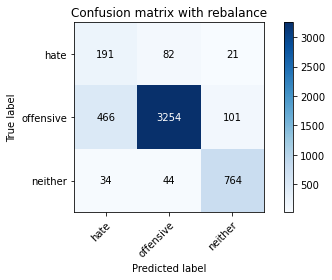
[ 34 44 764]]

Normalized confusion matrix

[[0.64965986 0.27891156 0.07142857]

[0.1219576 0.85160953 0.02643287]

[0.04038005 0.05225653 0.90736342]]





precision recall f1-score support

hate 0.28 0.65 0.39 294

offensive 0.96 0.85 0.90 3821

neither 0.86 0.91 0.88 842

accuracy 0.85 4957

macro avg 0.70 0.80 0.73 4957

weighted avg 0.90 0.85 0.87 4957

average eval\_loss: 0.43724961086897357, accuracy: 0.8491022796045996

training 5 epoch...

44.75324583053589 seconds used......

5 training finished! average train loss: 0.08881000982739075

total clf loss: 15.325226409826428 total adv loss: 2.774294198120515

evaluating...

Confusion matrix, without normalization

[[ 224 58 12]

[ 790 2940 91]

[ 62 39 741]]

Normalized confusion matrix

[[0.76190476 0.19727891 0.04081633]

[0.20675216 0.76943209 0.02381576]

[0.0736342 0.04631829 0.88004751]]





precision recall f1-score support

hate 0.21 0.76 0.33 294

offensive 0.97 0.77 0.86 3821

neither 0.88 0.88 0.88 842

accuracy 0.79 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.6893925027350986, accuracy: 0.7877748638289288

training 6 epoch...

44.810784578323364 seconds used......

6 training finished! average train loss: 0.033651440835977244

total clf loss: 5.801713918917812 total adv loss: 0.9463497422202636

evaluating...

Confusion matrix, without normalization

[[ 227 54 13]

[ 800 2929 92]

[ 77 41 724]]

Normalized confusion matrix

[[0.77210884 0.18367347 0.04421769]

[0.20936928 0.76655326 0.02407747]

[0.09144893 0.04869359 0.85985748]]





precision recall f1-score support

hate 0.21 0.77 0.32 294

offensive 0.97 0.77 0.86 3821

neither 0.87 0.86 0.87 842

accuracy 0.78 4957

macro avg 0.68 0.80 0.68 4957

weighted avg 0.91 0.78 0.83 4957

average eval\_loss: 0.8871386431921799, accuracy: 0.7827314908210611

training 7 epoch...

44.75350213050842 seconds used......

7 training finished! average train loss: 0.028719650031399657

total clf loss: 4.942234439193271 total adv loss: 0.6234848607316508

evaluating...

Confusion matrix, without normalization

[[ 238 42 14]

[ 925 2825 71]

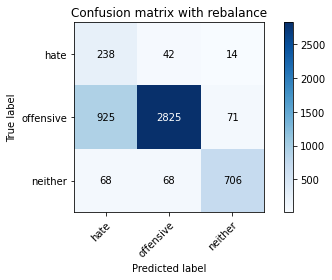
[ 68 68 706]]

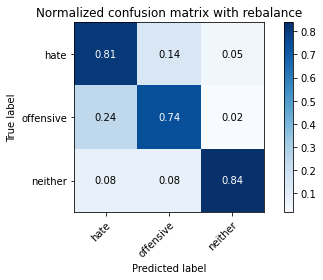
Normalized confusion matrix

[[0.80952381 0.14285714 0.04761905]

[0.24208322 0.73933525 0.01858152]

[0.0807601 0.0807601 0.83847981]]





precision recall f1-score support

hate 0.19 0.81 0.31 294

offensive 0.96 0.74 0.84 3821

neither 0.89 0.84 0.86 842

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: 1.070396041813036, accuracy: 0.7603389146661287

training 8 epoch...

44.817368507385254 seconds used......

8 training finished! average train loss: 0.027985077414985703

total clf loss: 4.832237985450774 total adv loss: 0.9357952354904455

evaluating...

Confusion matrix, without normalization

[[ 235 47 12]

[ 842 2916 63]

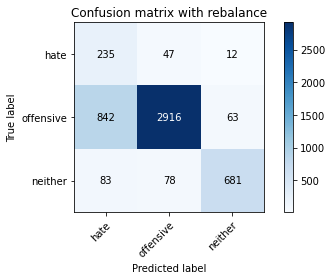
[ 83 78 681]]

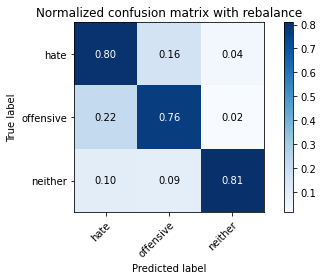
Normalized confusion matrix

[[0.79931973 0.15986395 0.04081633]

[0.22036116 0.76315101 0.01648783]

[0.09857482 0.09263658 0.8087886 ]]





precision recall f1-score support

hate 0.20 0.80 0.32 294

offensive 0.96 0.76 0.85 3821

neither 0.90 0.81 0.85 842

accuracy 0.77 4957

macro avg 0.69 0.79 0.68 4957

weighted avg 0.90 0.77 0.82 4957

average eval\_loss: 0.9620830061086905, accuracy: 0.7730482146459552

training 9 epoch...

44.89698314666748 seconds used......

9 training finished! average train loss: 0.023047180221026642

total clf loss: 4.0560177443840075 total adv loss: 2.298998431618429

evaluating...

Confusion matrix, without normalization

[[ 225 46 23]

[ 870 2817 134]

[ 49 20 773]]

Normalized confusion matrix

[[0.76530612 0.15646259 0.07823129]

[0.22768909 0.73724156 0.03506935]

[0.05819477 0.02375297 0.91805226]]





precision recall f1-score support

hate 0.20 0.77 0.31 294

offensive 0.98 0.74 0.84 3821

neither 0.83 0.92 0.87 842

accuracy 0.77 4957

macro avg 0.67 0.81 0.68 4957

weighted avg 0.91 0.77 0.81 4957

average eval\_loss: 1.3373959364271872, accuracy: 0.7696187210006052

training 10 epoch...

44.75310444831848 seconds used......

10 training finished! average train loss: 0.028484502184420567

total clf loss: 4.951081578474259 total adv loss: 1.6046338899980412

evaluating...

Confusion matrix, without normalization

[[ 210 64 20]

[ 638 3075 108]

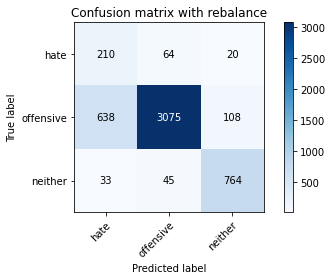
[ 33 45 764]]

Normalized confusion matrix

[[0.71428571 0.21768707 0.06802721]

[0.166972 0.80476315 0.02826485]

[0.0391924 0.05344418 0.90736342]]





precision recall f1-score support

hate 0.24 0.71 0.36 294

offensive 0.97 0.80 0.88 3821

neither 0.86 0.91 0.88 842

accuracy 0.82 4957

macro avg 0.69 0.81 0.71 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.8932532875372882, accuracy: 0.8168246923542465