hate as offensive

alpha=0.05

1 run...

Downloading: 100%

411/411 [00:00<00:00, 14.9kB/s]

Downloading: 100%

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

training 1 epoch...

45.13062381744385 seconds used......

1 training finished! average train loss: 0.6480458092968366

total clf loss: 113.67649871855974 total adv loss: 57.213304011151195

evaluating...

Confusion matrix, without normalization

[[ 233 26 35]

[1169 2509 143]

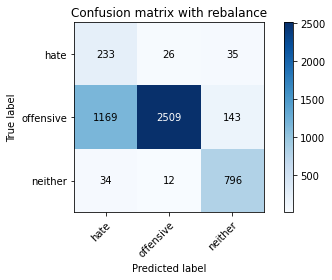
[ 34 12 796]]

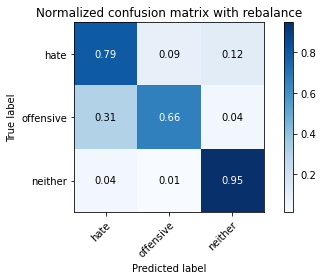
Normalized confusion matrix

[[0.79251701 0.08843537 0.11904762]

[0.30594085 0.65663439 0.03742476]

[0.04038005 0.01425178 0.94536817]]





precision recall f1-score support

hate 0.16 0.79 0.27 294

offensive 0.99 0.66 0.79 3821

neither 0.82 0.95 0.88 842

accuracy 0.71 4957

macro avg 0.65 0.80 0.64 4957

weighted avg 0.91 0.71 0.77 4957

average eval\_loss: 0.6996986309607183, accuracy: 0.7137381480734315

training 2 epoch...

44.58357095718384 seconds used......

2 training finished! average train loss: 0.40672281493394696

total clf loss: 70.9363628551364 total adv loss: 27.735227460041642

evaluating...

Confusion matrix, without normalization

[[ 210 69 15]

[ 432 3282 107]

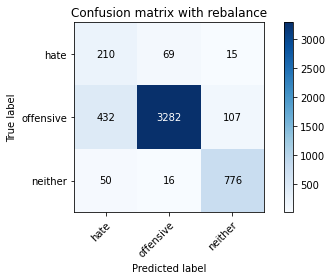
[ 50 16 776]]

Normalized confusion matrix

[[0.71428571 0.23469388 0.05102041]

[0.11305941 0.85893745 0.02800314]

[0.05938242 0.01900238 0.9216152 ]]





precision recall f1-score support

hate 0.30 0.71 0.43 294

offensive 0.97 0.86 0.91 3821

neither 0.86 0.92 0.89 842

accuracy 0.86 4957

macro avg 0.71 0.83 0.74 4957

weighted avg 0.92 0.86 0.88 4957

average eval\_loss: 0.40645463950932026, accuracy: 0.8610046399031672

training 3 epoch...

44.647796630859375 seconds used......

3 training finished! average train loss: 0.2795331673315394

total clf loss: 48.575854782015085 total adv loss: 15.513662069337443

evaluating...

Confusion matrix, without normalization

[[ 223 57 14]

[ 714 3000 107]

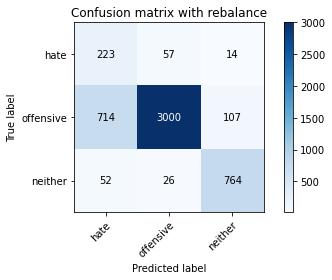
[ 52 26 764]]

Normalized confusion matrix

[[0.7585034 0.19387755 0.04761905]

[0.18686208 0.78513478 0.02800314]

[0.06175772 0.03087886 0.90736342]]





precision recall f1-score support

hate 0.23 0.76 0.35 294

offensive 0.97 0.79 0.87 3821

neither 0.86 0.91 0.88 842

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: 0.4824175441998147, accuracy: 0.8043171272947347

training 4 epoch...

44.676499366760254 seconds used......

4 training finished! average train loss: 0.1630550304329709

total clf loss: 28.133826149627566 total adv loss: 5.0283179615362315

evaluating...

Confusion matrix, without normalization

[[ 199 77 18]

[ 517 3202 102]

[ 44 50 748]]

Normalized confusion matrix

[[0.67687075 0.26190476 0.06122449]

[0.13530489 0.83800052 0.02669458]

[0.05225653 0.05938242 0.88836105]]





precision recall f1-score support

hate 0.26 0.68 0.38 294

offensive 0.96 0.84 0.90 3821

neither 0.86 0.89 0.87 842

accuracy 0.84 4957

macro avg 0.70 0.80 0.72 4957

weighted avg 0.90 0.84 0.86 4957

average eval\_loss: 0.4290235653414481, accuracy: 0.8369981843857172

training 5 epoch...

44.506911516189575 seconds used......

5 training finished! average train loss: 0.09442220027531274

total clf loss: 16.30443464871496 total adv loss: 3.16476807109575

evaluating...

Confusion matrix, without normalization

[[ 241 44 9]

[ 950 2783 88]

[ 121 22 699]]

Normalized confusion matrix

[[0.81972789 0.14965986 0.03061224]

[0.24862601 0.72834337 0.02303062]

[0.14370546 0.02612827 0.83016627]]





precision recall f1-score support

hate 0.18 0.82 0.30 294

offensive 0.98 0.73 0.83 3821

neither 0.88 0.83 0.85 842

accuracy 0.75 4957

macro avg 0.68 0.79 0.66 4957

weighted avg 0.91 0.75 0.81 4957

average eval\_loss: 0.9142549797712315, accuracy: 0.7510591083316522

training 6 epoch...

44.50615477561951 seconds used......

6 training finished! average train loss: 0.049092874203928424

total clf loss: 8.524083459400572 total adv loss: 2.5840398068667128

evaluating...

Confusion matrix, without normalization

[[ 205 59 30]

[ 666 3004 151]

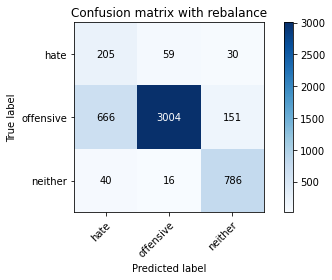
[ 40 16 786]]

Normalized confusion matrix

[[0.69727891 0.20068027 0.10204082]

[0.17429992 0.78618163 0.03951845]

[0.04750594 0.01900238 0.93349169]]





precision recall f1-score support

hate 0.23 0.70 0.34 294

offensive 0.98 0.79 0.87 3821

neither 0.81 0.93 0.87 842

accuracy 0.81 4957

macro avg 0.67 0.81 0.69 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.8298761573856727, accuracy: 0.8059310066572524

training 7 epoch...

44.5124671459198 seconds used......

7 training finished! average train loss: 0.027140290867156018

total clf loss: 4.691136092995293 total adv loss: 1.002927426975475

evaluating...

Confusion matrix, without normalization

[[ 209 63 22]

[ 615 3103 103]

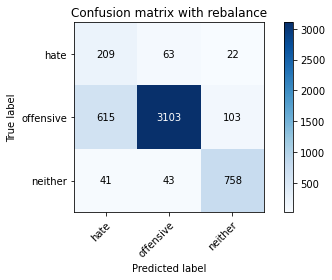
[ 41 43 758]]

Normalized confusion matrix

[[0.71088435 0.21428571 0.07482993]

[0.16095263 0.81209108 0.02695629]

[0.04869359 0.05106888 0.90023753]]





precision recall f1-score support

hate 0.24 0.71 0.36 294

offensive 0.97 0.81 0.88 3821

neither 0.86 0.90 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: 0.8267854071835303, accuracy: 0.8210611256808553

training 8 epoch...

44.55585026741028 seconds used......

8 training finished! average train loss: 0.0335155463139581

total clf loss: 5.825947341858409 total adv loss: 1.895778421082241

evaluating...

Confusion matrix, without normalization

[[ 201 78 15]

[ 481 3255 85]

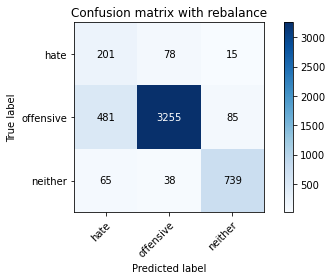
[ 65 38 739]]

Normalized confusion matrix

[[0.68367347 0.26530612 0.05102041]

[0.12588328 0.85187124 0.02224549]

[0.07719715 0.04513064 0.87767221]]





precision recall f1-score support

hate 0.27 0.68 0.39 294

offensive 0.97 0.85 0.91 3821

neither 0.88 0.88 0.88 842

accuracy 0.85 4957

macro avg 0.71 0.80 0.72 4957

weighted avg 0.91 0.85 0.87 4957

average eval\_loss: 0.7197467657181644, accuracy: 0.8462779907201937

training 9 epoch...

44.55278015136719 seconds used......

9 training finished! average train loss: 0.017486616828865563

total clf loss: 3.009025343169924 total adv loss: 0.37627716536634637

evaluating...

Confusion matrix, without normalization

[[ 226 51 17]

[ 735 2986 100]

[ 70 27 745]]

Normalized confusion matrix

[[0.76870748 0.17346939 0.05782313]

[0.19235802 0.78147082 0.02617116]

[0.08313539 0.03206651 0.8847981 ]]





precision recall f1-score support

hate 0.22 0.77 0.34 294

offensive 0.97 0.78 0.87 3821

neither 0.86 0.88 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: 1.0458934603187615, accuracy: 0.7982650796852935

training 10 epoch...

44.515077352523804 seconds used......

10 training finished! average train loss: 0.014036885896025019

total clf loss: 2.4012528533348814 total adv loss: 0.018906945886278947

evaluating...

Confusion matrix, without normalization

[[ 217 56 21]

[ 691 3009 121]

[ 45 28 769]]

Normalized confusion matrix

[[0.73809524 0.19047619 0.07142857]

[0.18084271 0.78749019 0.0316671 ]

[0.05344418 0.03325416 0.91330166]]





precision recall f1-score support

hate 0.23 0.74 0.35 294

offensive 0.97 0.79 0.87 3821

neither 0.84 0.91 0.88 842

accuracy 0.81 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 1.0594863955216902, accuracy: 0.8059310066572524

2 run...

training 1 epoch...

44.496989727020264 seconds used......

1 training finished! average train loss: 0.7367701160106045

total clf loss: 129.31111542880535 total adv loss: 66.46850589476526

evaluating...

Confusion matrix, without normalization

[[ 217 44 33]

[ 749 2938 134]

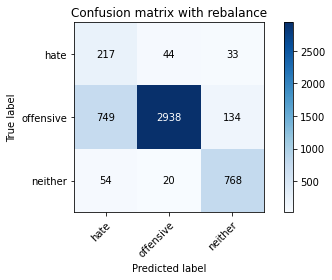
[ 54 20 768]]

Normalized confusion matrix

[[0.73809524 0.14965986 0.1122449 ]

[0.19602198 0.76890866 0.03506935]

[0.06413302 0.02375297 0.91211401]]





precision recall f1-score support

hate 0.21 0.74 0.33 294

offensive 0.98 0.77 0.86 3821

neither 0.82 0.91 0.86 842

accuracy 0.79 4957

macro avg 0.67 0.81 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.48343667988815614, accuracy: 0.7914060923945935

training 2 epoch...

44.456687211990356 seconds used......

2 training finished! average train loss: 0.40253456050192404

total clf loss: 70.2422196418047 total adv loss: 28.17619417142123

evaluating...

Confusion matrix, without normalization

[[ 227 44 23]

[ 700 2957 164]

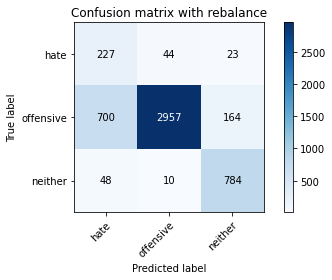
[ 48 10 784]]

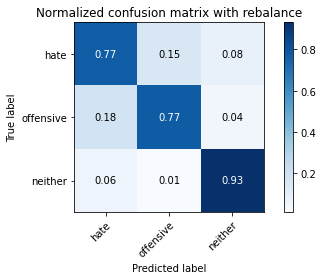
Normalized confusion matrix

[[0.77210884 0.14965986 0.07823129]

[0.18319812 0.77388118 0.0429207 ]

[0.05700713 0.01187648 0.93111639]]





precision recall f1-score support

hate 0.23 0.77 0.36 294

offensive 0.98 0.77 0.87 3821

neither 0.81 0.93 0.86 842

accuracy 0.80 4957

macro avg 0.67 0.83 0.70 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.5042402913733837, accuracy: 0.8004841638087553

training 3 epoch...

44.4827241897583 seconds used......

3 training finished! average train loss: 0.2760661596443221

total clf loss: 47.92974192649126 total adv loss: 14.448573750210926

evaluating...

Confusion matrix, without normalization

[[ 222 45 27]

[ 811 2855 155]

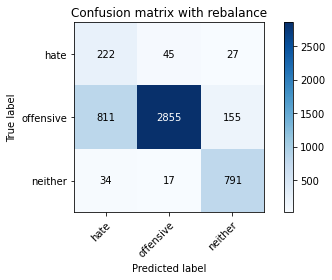
[ 34 17 791]]

Normalized confusion matrix

[[0.75510204 0.15306122 0.09183673]

[0.2122481 0.7471866 0.0405653 ]

[0.04038005 0.02019002 0.93942993]]





precision recall f1-score support

hate 0.21 0.76 0.33 294

offensive 0.98 0.75 0.85 3821

neither 0.81 0.94 0.87 842

accuracy 0.78 4957

macro avg 0.67 0.81 0.68 4957

weighted avg 0.90 0.78 0.82 4957

average eval\_loss: 0.5599648284185077, accuracy: 0.7803106717772846

training 4 epoch...

44.39658832550049 seconds used......

4 training finished! average train loss: 0.16058168543927503

total clf loss: 27.73630080744624 total adv loss: 5.536650192079833

evaluating...

Confusion matrix, without normalization

[[ 216 61 17]

[ 672 3040 109]

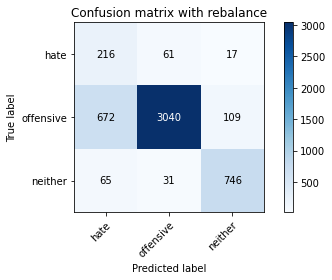
[ 65 31 746]]

Normalized confusion matrix

[[0.73469388 0.20748299 0.05782313]

[0.17587019 0.79560325 0.02852656]

[0.07719715 0.0368171 0.88598575]]





precision recall f1-score support

hate 0.23 0.73 0.35 294

offensive 0.97 0.80 0.87 3821

neither 0.86 0.89 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 0.552561684241218, accuracy: 0.8073431510994553

training 5 epoch...

44.42192029953003 seconds used......

5 training finished! average train loss: 0.09630975289248496

total clf loss: 16.621802440844476 total adv loss: 3.0566931800858583

evaluating...

Confusion matrix, without normalization

[[ 230 41 23]

[ 832 2839 150]

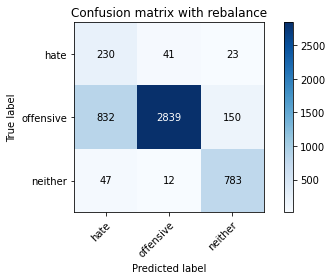
[ 47 12 783]]

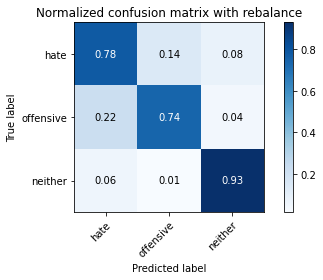
Normalized confusion matrix

[[0.78231293 0.13945578 0.07823129]

[0.21774405 0.74299921 0.03925674]

[0.05581948 0.01425178 0.92992874]]





precision recall f1-score support

hate 0.21 0.78 0.33 294

offensive 0.98 0.74 0.85 3821

neither 0.82 0.93 0.87 842

accuracy 0.78 4957

macro avg 0.67 0.82 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.799476096231956, accuracy: 0.7770829130522493

training 6 epoch...

44.3738055229187 seconds used......

6 training finished! average train loss: 0.048762254788654674

total clf loss: 8.404871630948037 total adv loss: 1.33052052732819

evaluating...

Confusion matrix, without normalization

[[ 219 60 15]

[ 806 2900 115]

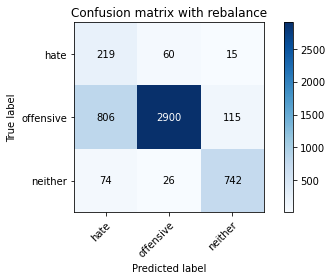
[ 74 26 742]]

Normalized confusion matrix

[[0.74489796 0.20408163 0.05102041]

[0.21093954 0.75896362 0.03009683]

[0.08788599 0.03087886 0.88123515]]





precision recall f1-score support

hate 0.20 0.74 0.31 294

offensive 0.97 0.76 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: 0.8843346090617424, accuracy: 0.7788985273350817

training 7 epoch...

44.37578010559082 seconds used......

7 training finished! average train loss: 0.013331253112365944

total clf loss: 2.297254893288482 total adv loss: 0.35221225011252955

evaluating...

Confusion matrix, without normalization

[[ 217 53 24]

[ 767 2910 144]

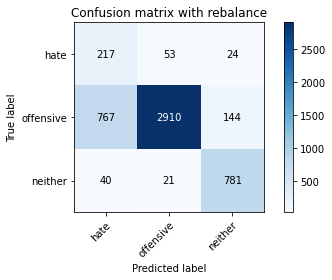
[ 40 21 781]]

Normalized confusion matrix

[[0.73809524 0.18027211 0.08163265]

[0.20073279 0.76158074 0.03768647]

[0.04750594 0.02494062 0.92755344]]





precision recall f1-score support

hate 0.21 0.74 0.33 294

offensive 0.98 0.76 0.86 3821

neither 0.82 0.93 0.87 842

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: 1.0475604247992798, accuracy: 0.7883800685898729

training 8 epoch...

44.481964349746704 seconds used......

8 training finished! average train loss: 0.0090393965298137

total clf loss: 1.5556921204261016 total adv loss: 0.19910629324155593

evaluating...

Confusion matrix, without normalization

[[ 217 53 24]

[ 760 2935 126]

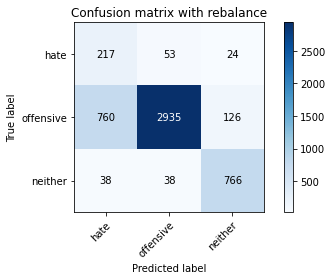
[ 38 38 766]]

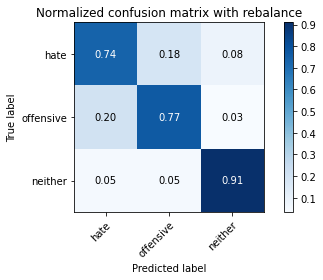
Normalized confusion matrix

[[0.73809524 0.18027211 0.08163265]

[0.19890081 0.76812353 0.03297566]

[0.04513064 0.04513064 0.90973872]]





precision recall f1-score support

hate 0.21 0.74 0.33 294

offensive 0.97 0.77 0.86 3821

neither 0.84 0.91 0.87 842

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: 1.1073610077134706, accuracy: 0.79039741779302

training 9 epoch...

44.4277617931366 seconds used......

9 training finished! average train loss: 0.030561202428506535

total clf loss: 5.37176976542105 total adv loss: 2.916082653330477

evaluating...

Confusion matrix, without normalization

[[ 210 61 23]

[ 618 3056 147]

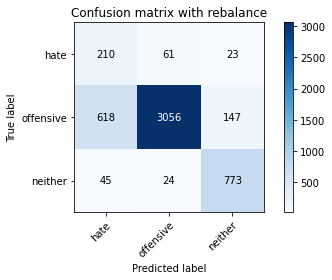
[ 45 24 773]]

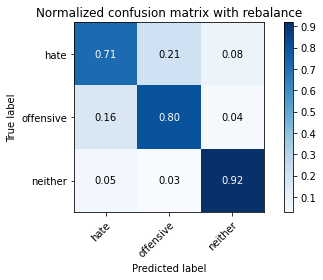
Normalized confusion matrix

[[0.71428571 0.20748299 0.07823129]

[0.16173776 0.79979063 0.0384716 ]

[0.05344418 0.02850356 0.91805226]]





precision recall f1-score support

hate 0.24 0.71 0.36 294

offensive 0.97 0.80 0.88 3821

neither 0.82 0.92 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.90 0.81 0.85 4957

average eval\_loss: 0.9310977342763855, accuracy: 0.8148073431510995

training 10 epoch...

44.43764305114746 seconds used......

10 training finished! average train loss: 0.014942047335544453

total clf loss: 2.6499173758202232 total adv loss: 1.8965455956572441

evaluating...

Confusion matrix, without normalization

[[ 195 68 31]

[ 502 3150 169]

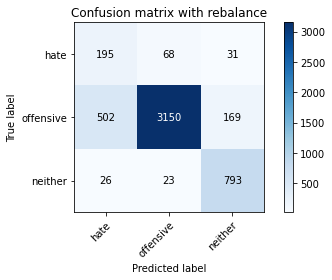
[ 26 23 793]]

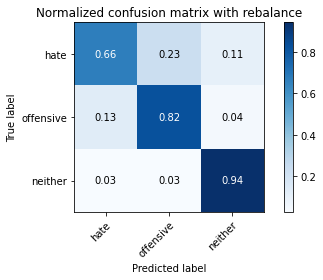
Normalized confusion matrix

[[0.66326531 0.23129252 0.10544218]

[0.13137922 0.82439152 0.04422926]

[0.03087886 0.02731591 0.94180523]]





precision recall f1-score support

hate 0.27 0.66 0.38 294

offensive 0.97 0.82 0.89 3821

neither 0.80 0.94 0.86 842

accuracy 0.83 4957

macro avg 0.68 0.81 0.71 4957

weighted avg 0.90 0.83 0.86 4957

average eval\_loss: 0.9141624850322372, accuracy: 0.8347791002622554

3 run...

training 1 epoch...

44.399261713027954 seconds used......

1 training finished! average train loss: 0.6414121972364292

total clf loss: 112.37872511148453 total adv loss: 53.944787826389074

evaluating...

Confusion matrix, without normalization

[[ 234 35 25]

[ 864 2857 100]

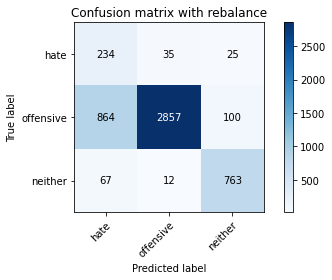
[ 67 12 763]]

Normalized confusion matrix

[[0.79591837 0.11904762 0.08503401]

[0.22611882 0.74771002 0.02617116]

[0.07957245 0.01425178 0.90617577]]





precision recall f1-score support

hate 0.20 0.80 0.32 294

offensive 0.98 0.75 0.85 3821

neither 0.86 0.91 0.88 842

accuracy 0.78 4957

macro avg 0.68 0.82 0.68 4957

weighted avg 0.92 0.78 0.82 4957

average eval\_loss: 0.5494627128204992, accuracy: 0.7774863828928787

training 2 epoch...

44.46412801742554 seconds used......

2 training finished! average train loss: 0.3881332083490857

total clf loss: 67.750927567482 total adv loss: 27.602978606708348

evaluating...

Confusion matrix, without normalization

[[ 244 31 19]

[1048 2666 107]

[ 51 20 771]]

Normalized confusion matrix

[[0.82993197 0.10544218 0.06462585]

[0.27427375 0.69772311 0.02800314]

[0.06057007 0.02375297 0.91567696]]





precision recall f1-score support

hate 0.18 0.83 0.30 294

offensive 0.98 0.70 0.82 3821

neither 0.86 0.92 0.89 842

accuracy 0.74 4957

macro avg 0.67 0.81 0.67 4957

weighted avg 0.91 0.74 0.80 4957

average eval\_loss: 0.5802014993443605, accuracy: 0.7425862416784346

training 3 epoch...

44.559818506240845 seconds used......

3 training finished! average train loss: 0.27030979377920167

total clf loss: 46.96665461733937 total adv loss: 14.873599346145056

evaluating...

Confusion matrix, without normalization

[[ 184 81 29]

[ 384 3298 139]

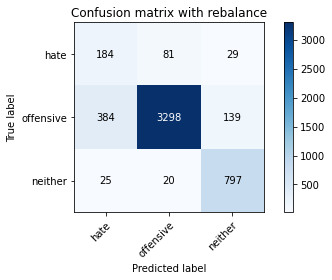
[ 25 20 797]]

Normalized confusion matrix

[[0.62585034 0.2755102 0.09863946]

[0.10049725 0.86312484 0.03637791]

[0.02969121 0.02375297 0.94655582]]





precision recall f1-score support

hate 0.31 0.63 0.41 294

offensive 0.97 0.86 0.91 3821

neither 0.83 0.95 0.88 842

accuracy 0.86 4957

macro avg 0.70 0.81 0.74 4957

weighted avg 0.91 0.86 0.88 4957

average eval\_loss: 0.38514002714486373, accuracy: 0.8632237240266291

training 4 epoch...

44.522000789642334 seconds used......

4 training finished! average train loss: 0.1422669033231268

total clf loss: 24.611047631129622 total adv loss: 5.668142217444256

evaluating...

Confusion matrix, without normalization

[[ 212 59 23]

[ 770 2898 153]

[ 45 15 782]]

Normalized confusion matrix

[[0.72108844 0.20068027 0.07823129]

[0.20151793 0.7584402 0.04004187]

[0.05344418 0.01781473 0.92874109]]





precision recall f1-score support

hate 0.21 0.72 0.32 294

offensive 0.98 0.76 0.85 3821

neither 0.82 0.93 0.87 842

accuracy 0.79 4957

macro avg 0.67 0.80 0.68 4957

weighted avg 0.90 0.79 0.82 4957

average eval\_loss: 0.6564946846901288, accuracy: 0.7851523098648376

training 5 epoch...

44.58471941947937 seconds used......

5 training finished! average train loss: 0.0808432012910775

total clf loss: 13.981460073962808 total adv loss: 3.1454518380178342

evaluating...

Confusion matrix, without normalization

[[ 243 36 15]

[1043 2680 98]

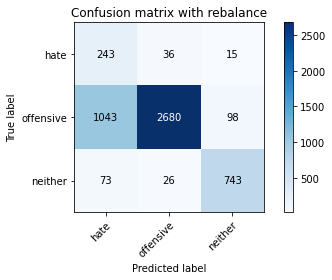
[ 73 26 743]]

Normalized confusion matrix

[[0.82653061 0.12244898 0.05102041]

[0.27296519 0.70138707 0.02564774]

[0.08669834 0.03087886 0.8824228 ]]





precision recall f1-score support

hate 0.18 0.83 0.29 294

offensive 0.98 0.70 0.82 3821

neither 0.87 0.88 0.88 842

accuracy 0.74 4957

macro avg 0.67 0.80 0.66 4957

weighted avg 0.91 0.74 0.80 4957

average eval\_loss: 0.8799471808522339, accuracy: 0.739560217873714

training 6 epoch...

44.6022412776947 seconds used......

6 training finished! average train loss: 0.03612011866385324

total clf loss: 6.261398160015233 total adv loss: 1.697158047513767

evaluating...

Confusion matrix, without normalization

[[ 208 69 17]

[ 531 3185 105]

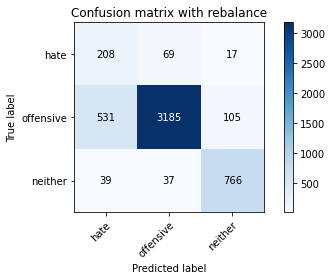
[ 39 37 766]]

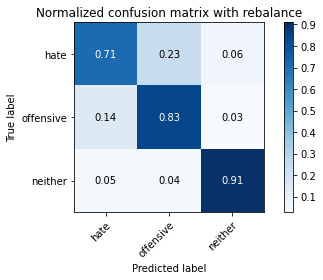
Normalized confusion matrix

[[0.70748299 0.23469388 0.05782313]

[0.13896886 0.83355143 0.02747972]

[0.04631829 0.04394299 0.90973872]]





precision recall f1-score support

hate 0.27 0.71 0.39 294

offensive 0.97 0.83 0.90 3821

neither 0.86 0.91 0.89 842

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: 0.6947124630481291, accuracy: 0.8390155335888643

training 7 epoch...

44.577064990997314 seconds used......

7 training finished! average train loss: 0.018841348470349344

total clf loss: 3.272678478213493 total adv loss: 1.0161579679588613

evaluating...

Confusion matrix, without normalization

[[ 234 49 11]

[ 830 2912 79]

[ 95 19 728]]

Normalized confusion matrix

[[0.79591837 0.16666667 0.03741497]

[0.21722062 0.76210416 0.02067522]

[0.1128266 0.02256532 0.86460808]]





precision recall f1-score support

hate 0.20 0.80 0.32 294

offensive 0.98 0.76 0.86 3821

neither 0.89 0.86 0.88 842

accuracy 0.78 4957

macro avg 0.69 0.81 0.69 4957

weighted avg 0.92 0.78 0.83 4957

average eval\_loss: 1.0827774478858638, accuracy: 0.7815210812991729

training 8 epoch...

44.576467752456665 seconds used......

8 training finished! average train loss: 0.026649332181679912

total clf loss: 4.660915678192396 total adv loss: 2.077597609681902

evaluating...

Confusion matrix, without normalization

[[ 216 51 27]

[ 757 2920 144]

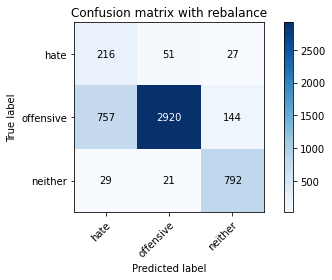
[ 29 21 792]]

Normalized confusion matrix

[[0.73469388 0.17346939 0.09183673]

[0.19811568 0.76419785 0.03768647]

[0.03444181 0.02494062 0.94061758]]





precision recall f1-score support

hate 0.22 0.73 0.33 294

offensive 0.98 0.76 0.86 3821

neither 0.82 0.94 0.88 842

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: 1.017681925392319, accuracy: 0.7924147669961671

training 9 epoch...

44.53172540664673 seconds used......

9 training finished! average train loss: 0.020065550590771364

total clf loss: 3.500301613181364 total adv loss: 1.3818494739798552

evaluating...

Confusion matrix, without normalization

[[ 218 58 18]

[ 739 2989 93]

[ 48 42 752]]

Normalized confusion matrix

[[0.7414966 0.19727891 0.06122449]

[0.19340487 0.78225595 0.02433918]

[0.05700713 0.04988124 0.89311164]]





precision recall f1-score support

hate 0.22 0.74 0.34 294

offensive 0.97 0.78 0.87 3821

neither 0.87 0.89 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: 0.9865274818636657, accuracy: 0.7986685495259229

training 10 epoch...

44.531898975372314 seconds used......

10 training finished! average train loss: 0.013058810734053208

total clf loss: 2.2551296504825586 total adv loss: 0.44146000918453865

evaluating...

Confusion matrix, without normalization

[[ 242 41 11]

[ 910 2812 99]

[ 82 23 737]]

Normalized confusion matrix

[[0.82312925 0.13945578 0.03741497]

[0.23815755 0.735933 0.02590945]

[0.09738717 0.02731591 0.87529691]]





precision recall f1-score support

hate 0.20 0.82 0.32 294

offensive 0.98 0.74 0.84 3821

neither 0.87 0.88 0.87 842

accuracy 0.76 4957

macro avg 0.68 0.81 0.68 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 1.417897031668784, accuracy: 0.7647770829130522

4 run...

training 1 epoch...

44.56849241256714 seconds used......

1 training finished! average train loss: 0.6145780496081413

total clf loss: 107.70105338096619 total adv loss: 52.164139145985246

evaluating...

Confusion matrix, without normalization

[[ 212 58 24]

[ 503 3176 142]

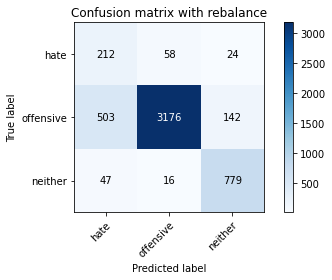
[ 47 16 779]]

Normalized confusion matrix

[[0.72108844 0.19727891 0.08163265]

[0.13164093 0.83119602 0.03716305]

[0.05581948 0.01900238 0.92517815]]





precision recall f1-score support

hate 0.28 0.72 0.40 294

offensive 0.98 0.83 0.90 3821

neither 0.82 0.93 0.87 842

accuracy 0.84 4957

macro avg 0.69 0.83 0.72 4957

weighted avg 0.91 0.84 0.86 4957

average eval\_loss: 0.37483258479304854, accuracy: 0.8406294129513819

training 2 epoch...

44.522265911102295 seconds used......

2 training finished! average train loss: 0.37269939804634855

total clf loss: 65.04219200462103 total adv loss: 26.211902770213783

evaluating...

Confusion matrix, without normalization

[[ 233 47 14]

[ 779 2940 102]

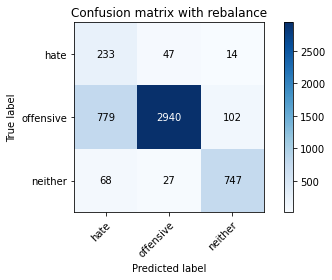
[ 68 27 747]]

Normalized confusion matrix

[[0.79251701 0.15986395 0.04761905]

[0.20387333 0.76943209 0.02669458]

[0.0807601 0.03206651 0.8871734 ]]





precision recall f1-score support

hate 0.22 0.79 0.34 294

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: 0.5048354311335471, accuracy: 0.7908008876336494

training 3 epoch...

44.55632185935974 seconds used......

3 training finished! average train loss: 0.24464517202206523

total clf loss: 42.525178235024214 total adv loss: 13.817075430240948

evaluating...

Confusion matrix, without normalization

[[ 235 45 14]

[ 783 2923 115]

[ 55 17 770]]

Normalized confusion matrix

[[0.79931973 0.15306122 0.04761905]

[0.20492018 0.76498299 0.03009683]

[0.06532067 0.02019002 0.91448931]]





precision recall f1-score support

hate 0.22 0.80 0.34 294

offensive 0.98 0.76 0.86 3821

neither 0.86 0.91 0.88 842

accuracy 0.79 4957

macro avg 0.68 0.83 0.70 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.5653197021944628, accuracy: 0.7924147669961671

training 4 epoch...

44.565749168395996 seconds used......

4 training finished! average train loss: 0.13837340188740987

total clf loss: 23.97402056120336 total adv loss: 6.2433769768394995

evaluating...

Confusion matrix, without normalization

[[ 202 71 21]

[ 490 3194 137]

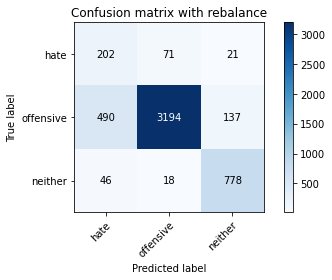
[ 46 18 778]]

Normalized confusion matrix

[[0.68707483 0.2414966 0.07142857]

[0.12823868 0.83590683 0.03585449]

[0.05463183 0.02137767 0.9239905 ]]





precision recall f1-score support

hate 0.27 0.69 0.39 294

offensive 0.97 0.84 0.90 3821

neither 0.83 0.92 0.88 842

accuracy 0.84 4957

macro avg 0.69 0.82 0.72 4957

weighted avg 0.91 0.84 0.87 4957

average eval\_loss: 0.5214820436217011, accuracy: 0.8420415573935849

training 5 epoch...

44.56461191177368 seconds used......

5 training finished! average train loss: 0.06187128625014982

total clf loss: 10.651368329767138 total adv loss: 1.427566732672858

evaluating...

Confusion matrix, without normalization

[[ 205 65 24]

[ 483 3201 137]

[ 33 22 787]]

Normalized confusion matrix

[[0.69727891 0.22108844 0.08163265]

[0.1264067 0.83773881 0.03585449]

[0.0391924 0.02612827 0.93467933]]





precision recall f1-score support

hate 0.28 0.70 0.40 294

offensive 0.97 0.84 0.90 3821

neither 0.83 0.93 0.88 842

accuracy 0.85 4957

macro avg 0.70 0.82 0.73 4957

weighted avg 0.91 0.85 0.87 4957

average eval\_loss: 0.5995940288079099, accuracy: 0.8458745208795643

training 6 epoch...

44.55971932411194 seconds used......

6 training finished! average train loss: 0.02843846762430315

total clf loss: 4.927490377915092 total adv loss: 1.2902490046799358

evaluating...

Confusion matrix, without normalization

[[ 238 40 16]

[ 987 2709 125]

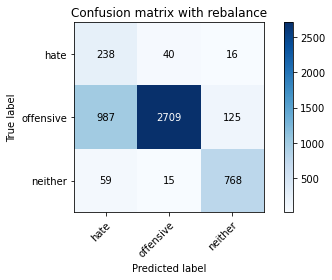
[ 59 15 768]]

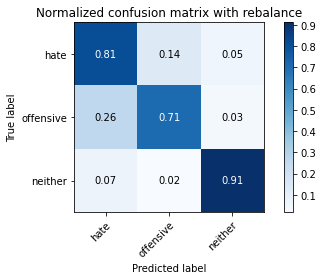
Normalized confusion matrix

[[0.80952381 0.13605442 0.05442177]

[0.25830934 0.70897671 0.03271395]

[0.07007126 0.01781473 0.91211401]]





precision recall f1-score support

hate 0.19 0.81 0.30 294

offensive 0.98 0.71 0.82 3821

neither 0.84 0.91 0.88 842

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: 1.190013982900118, accuracy: 0.7494452289691346

training 7 epoch...

44.53508400917053 seconds used......

7 training finished! average train loss: 0.022624539356230763

total clf loss: 3.9170857190911192 total adv loss: 0.9657894682659389

evaluating...

Confusion matrix, without normalization

[[ 240 35 19]

[ 984 2694 143]

[ 68 9 765]]

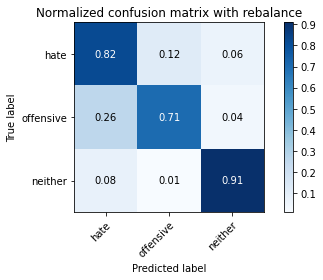
Normalized confusion matrix

[[0.81632653 0.11904762 0.06462585]

[0.25752421 0.70505103 0.03742476]

[0.0807601 0.01068884 0.90855107]]





precision recall f1-score support

hate 0.19 0.82 0.30 294

offensive 0.98 0.71 0.82 3821

neither 0.83 0.91 0.86 842

accuracy 0.75 4957

macro avg 0.66 0.81 0.66 4957

weighted avg 0.91 0.75 0.80 4957

average eval\_loss: 1.3353338527020007, accuracy: 0.7462174702440992

training 8 epoch...

44.5313560962677 seconds used......

8 training finished! average train loss: 0.03194576808730037

total clf loss: 5.548320832604077 total adv loss: 1.7118902186084597

evaluating...

Confusion matrix, without normalization

[[ 236 44 14]

[ 805 2934 82]

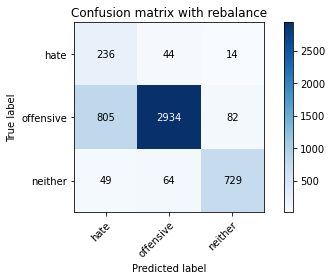
[ 49 64 729]]

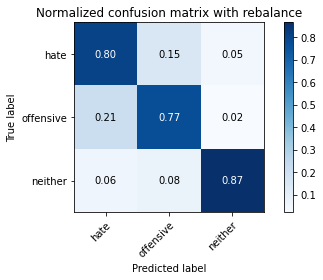
Normalized confusion matrix

[[0.80272109 0.14965986 0.04761905]

[0.21067783 0.76786182 0.02146035]

[0.05819477 0.0760095 0.86579572]]





precision recall f1-score support

hate 0.22 0.80 0.34 294

offensive 0.96 0.77 0.86 3821

neither 0.88 0.87 0.87 842

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: 1.0287378522062358, accuracy: 0.7865644543070406

training 9 epoch...

44.52340078353882 seconds used......

9 training finished! average train loss: 0.018834538388641943

total clf loss: 3.2860881939704996 total adv loss: 1.3076425998151535

evaluating...

Confusion matrix, without normalization

[[ 227 44 23]

[ 795 2861 165]

[ 36 15 791]]

Normalized confusion matrix

[[0.77210884 0.14965986 0.07823129]

[0.20806072 0.74875687 0.04318241]

[0.04275534 0.01781473 0.93942993]]





precision recall f1-score support

hate 0.21 0.77 0.34 294

offensive 0.98 0.75 0.85 3821

neither 0.81 0.94 0.87 842

accuracy 0.78 4957

macro avg 0.67 0.82 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 1.2037469289179201, accuracy: 0.7825297559007465

training 10 epoch...

44.52005481719971 seconds used......

10 training finished! average train loss: 0.010755670068163305

total clf loss: 1.8616508092964068 total adv loss: 0.4486245779970659

evaluating...

Confusion matrix, without normalization

[[ 203 60 31]

[ 578 3068 175]

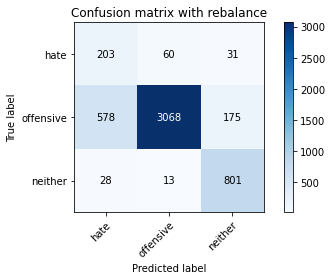
[ 28 13 801]]

Normalized confusion matrix

[[0.69047619 0.20408163 0.10544218]

[0.1512693 0.80293117 0.04579953]

[0.03325416 0.01543943 0.95130641]]





precision recall f1-score support

hate 0.25 0.69 0.37 294

offensive 0.98 0.80 0.88 3821

neither 0.80 0.95 0.87 842

accuracy 0.82 4957

macro avg 0.67 0.81 0.71 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.9325432035402628, accuracy: 0.8214645955214848