alpha = 1

theta=1

training 1 epoch...

61.780503273010254 seconds used......

1 training finished! average train loss: -2.6309853942073578

total clf loss: 129.64525672793388 total adv loss: 77.94012298434973 657.4838827252388

evaluating...

Confusion matrix, without normalization

[[ 227 41 26]

[ 818 2893 110]

[ 74 19 749]]

Normalized confusion matrix

[[0.77210884 0.13945578 0.08843537]

[0.21408008 0.75713164 0.02878828]

[0.08788599 0.02256532 0.88954869]]





precision recall f1-score support

hate 0.20 0.77 0.32 294

offensive 0.98 0.76 0.85 3821

neither 0.85 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: 0.5437022997366805, accuracy: 0.7805124066975994

training 2 epoch...

61.645538091659546 seconds used......

2 training finished! average train loss: -3.7792067388344925

total clf loss: 84.66922006011009 total adv loss: 57.58269960060716 788.4962737560272

evaluating...

Confusion matrix, without normalization

[[ 245 37 12]

[ 899 2870 52]

[ 201 12 629]]

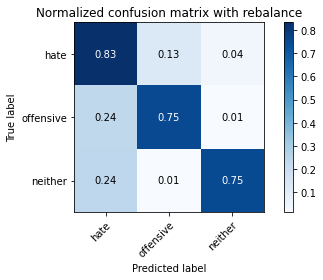
Normalized confusion matrix

[[0.83333333 0.12585034 0.04081633]

[0.23527872 0.75111227 0.013609 ]

[0.23871734 0.01425178 0.74703088]]





precision recall f1-score support

hate 0.18 0.83 0.30 294

offensive 0.98 0.75 0.85 3821

neither 0.91 0.75 0.82 842

accuracy 0.76 4957

macro avg 0.69 0.78 0.66 4957

weighted avg 0.92 0.76 0.81 4957

average eval\_loss: 0.5578454797787051, accuracy: 0.7552955416582611

training 3 epoch...

61.649054288864136 seconds used......

3 training finished! average train loss: -3.965661755779333

total clf loss: 67.7512726932764 total adv loss: 50.647318474948406 796.5267496109009

evaluating...

Confusion matrix, without normalization

[[ 233 49 12]

[ 684 3060 77]

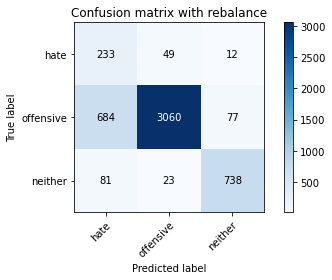
[ 81 23 738]]

Normalized confusion matrix

[[0.79251701 0.16666667 0.04081633]

[0.17901073 0.80083748 0.02015179]

[0.09619952 0.02731591 0.87648456]]





precision recall f1-score support

hate 0.23 0.79 0.36 294

offensive 0.98 0.80 0.88 3821

neither 0.89 0.88 0.88 842

accuracy 0.81 4957

macro avg 0.70 0.82 0.71 4957

weighted avg 0.92 0.81 0.85 4957

average eval\_loss: 0.45710281003266573, accuracy: 0.8131934637885818

training 4 epoch...

61.80384182929993 seconds used......

4 training finished! average train loss: -4.1332801754711666

total clf loss: 54.233543902635574 total adv loss: 42.08819995820522 803.1126537322998

evaluating...

Confusion matrix, without normalization

[[ 226 51 17]

[ 695 3014 112]

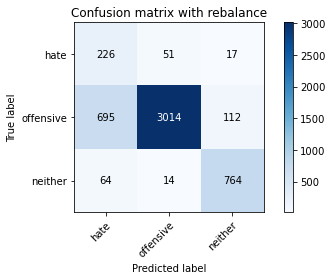
[ 64 14 764]]

Normalized confusion matrix

[[0.76870748 0.17346939 0.05782313]

[0.18188956 0.78879874 0.0293117 ]

[0.0760095 0.01662708 0.90736342]]





precision recall f1-score support

hate 0.23 0.77 0.35 294

offensive 0.98 0.79 0.87 3821

neither 0.86 0.91 0.88 842

accuracy 0.81 4957

macro avg 0.69 0.82 0.70 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 0.5027622977692273, accuracy: 0.8077466209400848

training 5 epoch...

61.86584138870239 seconds used......

5 training finished! average train loss: -5.585073185245893

total clf loss: 46.72542083263397 total adv loss: 31.971524389460683 1033.7444577217102

evaluating...

Confusion matrix, without normalization

[[ 203 72 19]

[ 498 3211 112]

[ 46 59 737]]

Normalized confusion matrix

[[0.69047619 0.24489796 0.06462585]

[0.13033237 0.84035593 0.0293117 ]

[0.05463183 0.07007126 0.87529691]]





precision recall f1-score support

hate 0.27 0.69 0.39 294

offensive 0.96 0.84 0.90 3821

neither 0.85 0.88 0.86 842

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: 0.4229572881013155, accuracy: 0.8374016542263466

training 6 epoch...

61.59804177284241 seconds used......

6 training finished! average train loss: -6.388365416498909

total clf loss: 40.7184327468276 total adv loss: 23.628373362123966 1156.757291316986

evaluating...

Confusion matrix, without normalization

[[ 222 54 18]

[ 815 2921 85]

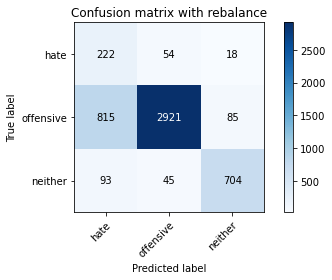
[ 93 45 704]]

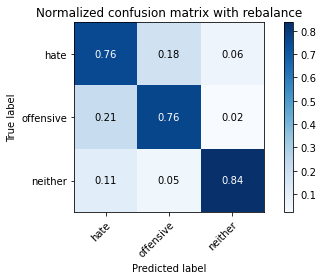
Normalized confusion matrix

[[0.75510204 0.18367347 0.06122449]

[0.21329495 0.76445957 0.02224549]

[0.11045131 0.05344418 0.83610451]]





precision recall f1-score support

hate 0.20 0.76 0.31 294

offensive 0.97 0.76 0.85 3821

neither 0.87 0.84 0.85 842

accuracy 0.78 4957

macro avg 0.68 0.79 0.67 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.5911231948543461, accuracy: 0.7760742384506758

training 7 epoch...

61.55777955055237 seconds used......

7 training finished! average train loss: -6.441481004681504

total clf loss: 37.44774106517434 total adv loss: 25.188385575078428 1164.1293768882751

evaluating...

Confusion matrix, without normalization

[[ 232 49 13]

[ 784 2946 91]

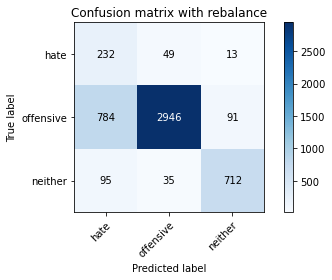
[ 95 35 712]]

Normalized confusion matrix

[[0.78911565 0.16666667 0.04421769]

[0.20518189 0.77100236 0.02381576]

[0.1128266 0.0415677 0.8456057 ]]





precision recall f1-score support

hate 0.21 0.79 0.33 294

offensive 0.97 0.77 0.86 3821

neither 0.87 0.85 0.86 842

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: 0.5711696122382437, accuracy: 0.7847488400242082

training 8 epoch...

61.48265862464905 seconds used......

8 training finished! average train loss: -6.693264133051822

total clf loss: 27.329474676400423 total adv loss: 13.840888824313879 1185.7185292243958

evaluating...

Confusion matrix, without normalization

[[ 191 72 31]

[ 477 3187 157]

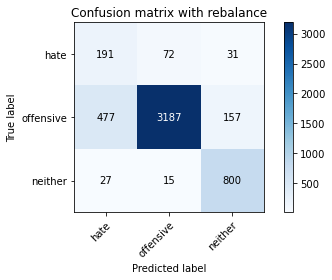
[ 27 15 800]]

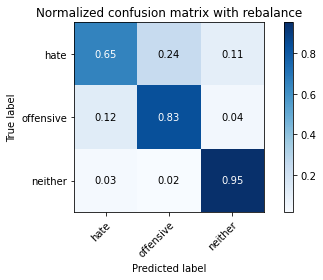
Normalized confusion matrix

[[0.64965986 0.24489796 0.10544218]

[0.12483643 0.83407485 0.04108872]

[0.03206651 0.01781473 0.95011876]]





precision recall f1-score support

hate 0.27 0.65 0.39 294

offensive 0.97 0.83 0.90 3821

neither 0.81 0.95 0.87 842

accuracy 0.84 4957

macro avg 0.69 0.81 0.72 4957

weighted avg 0.90 0.84 0.86 4957

average eval\_loss: 0.4412861344734988, accuracy: 0.8428484970748437

training 9 epoch...

61.48211359977722 seconds used......

9 training finished! average train loss: -6.717047415281597

total clf loss: 25.33576587215066 total adv loss: 11.406895129941404 1185.3577637672424

evaluating...

Confusion matrix, without normalization

[[ 219 62 13]

[ 660 3068 93]

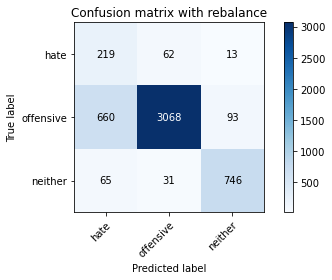
[ 65 31 746]]

Normalized confusion matrix

[[0.74489796 0.21088435 0.04421769]

[0.17272965 0.80293117 0.02433918]

[0.07719715 0.0368171 0.88598575]]





precision recall f1-score support

hate 0.23 0.74 0.35 294

offensive 0.97 0.80 0.88 3821

neither 0.88 0.89 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: 0.5156020419071278, accuracy: 0.8135969336292113

training 10 epoch...

61.50417876243591 seconds used......

10 training finished! average train loss: -6.896015808596249

total clf loss: 18.49301852285862 total adv loss: 6.868211300927214 1204.5799345970154

evaluating...

Confusion matrix, without normalization

[[ 221 59 14]

[ 680 3037 104]

[ 59 28 755]]

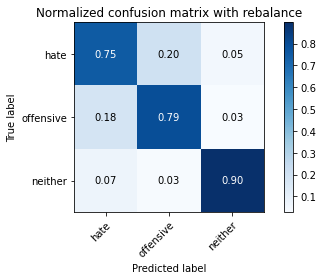
Normalized confusion matrix

[[0.75170068 0.20068027 0.04761905]

[0.17796388 0.79481811 0.02721801]

[0.07007126 0.03325416 0.89667458]]





precision recall f1-score support

hate 0.23 0.75 0.35 294

offensive 0.97 0.79 0.87 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 0.84 4957

average eval\_loss: 0.538868589401846, accuracy: 0.809562235222917

2 run

training 1 epoch...

62.10349178314209 seconds used......

1 training finished! average train loss: -2.114141183289868

total clf loss: 124.83115497231483 total adv loss: 78.76685675233603 565.1161540746689

evaluating...

Confusion matrix, without normalization

[[ 206 42 46]

[ 747 2853 221]

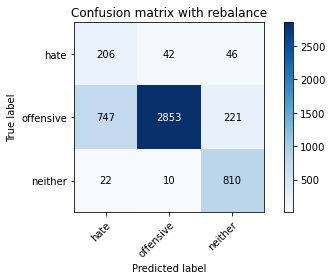
[ 22 10 810]]

Normalized confusion matrix

[[0.70068027 0.14285714 0.15646259]

[0.19549856 0.74666318 0.05783826]

[0.02612827 0.01187648 0.96199525]]





precision recall f1-score support

hate 0.21 0.70 0.32 294

offensive 0.98 0.75 0.85 3821

neither 0.75 0.96 0.84 842

accuracy 0.78 4957

macro avg 0.65 0.80 0.67 4957

weighted avg 0.90 0.78 0.82 4957

average eval\_loss: 0.6015703421566756, accuracy: 0.7805124066975994

training 2 epoch...

61.67193651199341 seconds used......

2 training finished! average train loss: -3.3954082350981865

total clf loss: 82.7190777361393 total adv loss: 57.87683969736099 721.2107262611389

evaluating...

Confusion matrix, without normalization

[[ 185 95 14]

[ 254 3499 68]

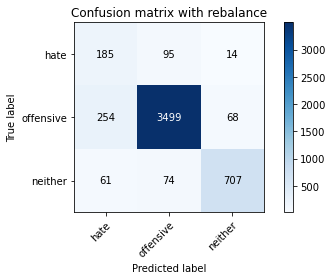
[ 61 74 707]]

Normalized confusion matrix

[[0.6292517 0.32312925 0.04761905]

[0.06647474 0.91572887 0.01779639]

[0.07244656 0.08788599 0.83966746]]





precision recall f1-score support

hate 0.37 0.63 0.47 294

offensive 0.95 0.92 0.93 3821

neither 0.90 0.84 0.87 842

accuracy 0.89 4957

macro avg 0.74 0.79 0.76 4957

weighted avg 0.91 0.89 0.90 4957

average eval\_loss: 0.37040299090046075, accuracy: 0.8858180351018762

training 3 epoch...

61.61112570762634 seconds used......

3 training finished! average train loss: -3.601083755493164

total clf loss: 65.48257206380367 total adv loss: 48.36739535816014 729.6352894306183

evaluating...

Confusion matrix, without normalization

[[ 226 53 15]

[ 716 2976 129]

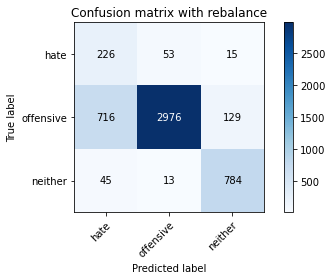
[ 45 13 784]]

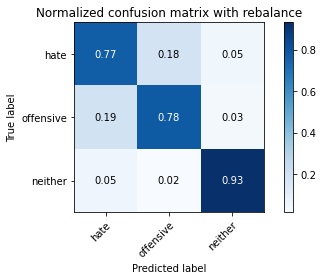
Normalized confusion matrix

[[0.76870748 0.18027211 0.05102041]

[0.1873855 0.7788537 0.0337608 ]

[0.05344418 0.01543943 0.93111639]]





precision recall f1-score support

hate 0.23 0.77 0.35 294

offensive 0.98 0.78 0.87 3821

neither 0.84 0.93 0.89 842

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: 0.47280012365550766, accuracy: 0.80411539237442

training 4 epoch...

61.6728196144104 seconds used......

4 training finished! average train loss: -4.034866623014038

total clf loss: 52.53065376728773 total adv loss: 39.94547991268337 782.4383244514465

evaluating...

Confusion matrix, without normalization

[[ 226 53 15]

[ 665 3064 92]

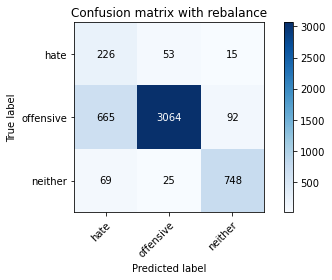
[ 69 25 748]]

Normalized confusion matrix

[[0.76870748 0.18027211 0.05102041]

[0.17403821 0.80188432 0.02407747]

[0.08194774 0.02969121 0.88836105]]





precision recall f1-score support

hate 0.24 0.77 0.36 294

offensive 0.98 0.80 0.88 3821

neither 0.87 0.89 0.88 842

accuracy 0.81 4957

macro avg 0.70 0.82 0.71 4957

weighted avg 0.91 0.81 0.85 4957

average eval\_loss: 0.5109099284895966, accuracy: 0.8146056082307848

training 5 epoch...

61.71439862251282 seconds used......

5 training finished! average train loss: -5.755475023336578

total clf loss: 50.17818791419268 total adv loss: 37.900113098323345 1072.2645292282104

evaluating...

Confusion matrix, without normalization

[[ 155 124 15]

[ 239 3506 76]

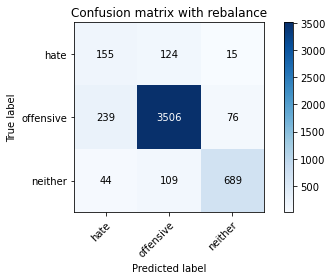
[ 44 109 689]]

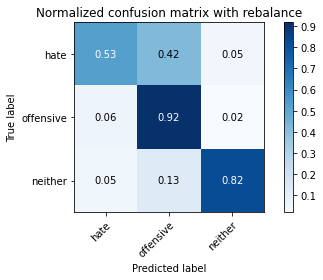
Normalized confusion matrix

[[0.52721088 0.42176871 0.05102041]

[0.06254907 0.91756085 0.01989008]

[0.05225653 0.12945368 0.81828979]]





precision recall f1-score support

hate 0.35 0.53 0.42 294

offensive 0.94 0.92 0.93 3821

neither 0.88 0.82 0.85 842

accuracy 0.88 4957

macro avg 0.72 0.75 0.73 4957

weighted avg 0.89 0.88 0.88 4957

average eval\_loss: 0.33811440375903923, accuracy: 0.8775469033689731

training 6 epoch...

61.584776639938354 seconds used......

6 training finished! average train loss: -5.932573340789617

total clf loss: 43.40113817900419 total adv loss: 27.817520085722208 1085.6886992454529

evaluating...

Confusion matrix, without normalization

[[ 232 42 20]

[ 791 2914 116]

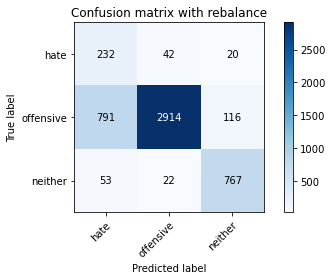
[ 53 22 767]]

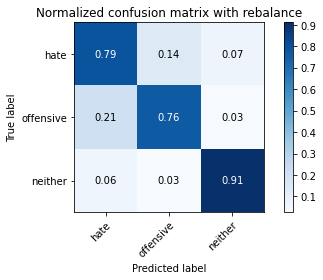
Normalized confusion matrix

[[0.78911565 0.14285714 0.06802721]

[0.20701387 0.76262758 0.03035854]

[0.06294537 0.02612827 0.91092637]]





precision recall f1-score support

hate 0.22 0.79 0.34 294

offensive 0.98 0.76 0.86 3821

neither 0.85 0.91 0.88 842

accuracy 0.79 4957

macro avg 0.68 0.82 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.5576160658271082, accuracy: 0.7893887431914465

training 7 epoch...

61.52955627441406 seconds used......

7 training finished! average train loss: -6.107172977157504

total clf loss: 35.40723202750087 total adv loss: 21.087543516419828 1100.8213543891907

evaluating...

Confusion matrix, without normalization

[[ 245 41 8]

[ 955 2794 72]

[ 117 47 678]]

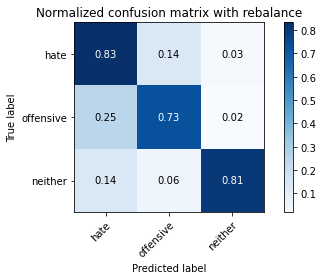
Normalized confusion matrix

[[0.83333333 0.13945578 0.02721088]

[0.24993457 0.73122219 0.01884323]

[0.13895487 0.05581948 0.80522565]]





precision recall f1-score support

hate 0.19 0.83 0.30 294

offensive 0.97 0.73 0.83 3821

neither 0.89 0.81 0.85 842

accuracy 0.75 4957

macro avg 0.68 0.79 0.66 4957

weighted avg 0.91 0.75 0.80 4957

average eval\_loss: 0.634600987447606, accuracy: 0.749848698809764

training 8 epoch...

61.474059104919434 seconds used......

8 training finished! average train loss: -6.203800477479634

total clf loss: 30.761011108756065 total adv loss: 17.02873453963548 1108.6396279335022

evaluating...

Confusion matrix, without normalization

[[ 216 70 8]

[ 668 3115 38]

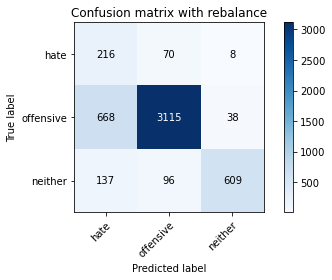
[ 137 96 609]]

Normalized confusion matrix

[[0.73469388 0.23809524 0.02721088]

[0.17482334 0.81523161 0.00994504]

[0.16270784 0.11401425 0.72327791]]





precision recall f1-score support

hate 0.21 0.73 0.33 294

offensive 0.95 0.82 0.88 3821

neither 0.93 0.72 0.81 842

accuracy 0.79 4957

macro avg 0.70 0.76 0.67 4957

weighted avg 0.90 0.79 0.83 4957

average eval\_loss: 0.5285874459171488, accuracy: 0.7948355860399435

training 9 epoch...

61.54219079017639 seconds used......

9 training finished! average train loss: -6.307674700753731

total clf loss: 25.519747633486986 total adv loss: 12.875327417394146 1117.0074524879456

evaluating...

Confusion matrix, without normalization

[[ 182 92 20]

[ 428 3269 124]

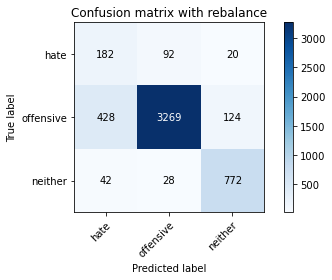
[ 42 28 772]]

Normalized confusion matrix

[[0.61904762 0.31292517 0.06802721]

[0.11201256 0.8555352 0.03245224]

[0.04988124 0.03325416 0.91686461]]





precision recall f1-score support

hate 0.28 0.62 0.38 294

offensive 0.96 0.86 0.91 3821

neither 0.84 0.92 0.88 842

accuracy 0.85 4957

macro avg 0.70 0.80 0.72 4957

weighted avg 0.90 0.85 0.87 4957

average eval\_loss: 0.41943324974647933, accuracy: 0.8519265684890055

training 10 epoch...

61.48135828971863 seconds used......

10 training finished! average train loss: -6.492569432621114

total clf loss: 17.26882349513471 total adv loss: 6.539132221485488 1134.0373306274414

evaluating...

Confusion matrix, without normalization

[[ 235 44 15]

[ 948 2722 151]

[ 62 31 749]]

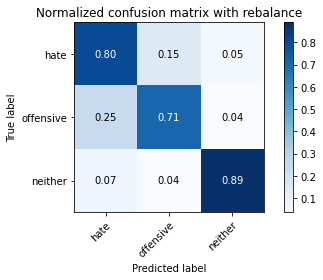
Normalized confusion matrix

[[0.79931973 0.14965986 0.05102041]

[0.24810259 0.71237896 0.03951845]

[0.0736342 0.0368171 0.88954869]]





precision recall f1-score support

hate 0.19 0.80 0.31 294

offensive 0.97 0.71 0.82 3821

neither 0.82 0.89 0.85 842

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: 0.6934944077004348, accuracy: 0.7476296146863022

3 run

training 1 epoch...

61.581631660461426 seconds used......

1 training finished! average train loss: -2.0835462432158622

total clf loss: 121.39044637978077 total adv loss: 76.57869404554367 554.2555482387543

evaluating...

Confusion matrix, without normalization

[[ 189 54 51]

[ 661 2944 216]

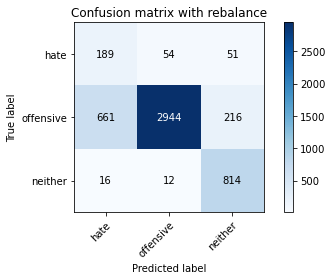
[ 16 12 814]]

Normalized confusion matrix

[[0.64285714 0.18367347 0.17346939]

[0.17299136 0.77047893 0.0565297 ]

[0.01900238 0.01425178 0.96674584]]





precision recall f1-score support

hate 0.22 0.64 0.33 294

offensive 0.98 0.77 0.86 3821

neither 0.75 0.97 0.85 842

accuracy 0.80 4957

macro avg 0.65 0.79 0.68 4957

weighted avg 0.89 0.80 0.83 4957

average eval\_loss: 0.4754839383306042, accuracy: 0.7962477304821465

training 2 epoch...

61.656139850616455 seconds used......

2 training finished! average train loss: -3.5805034079746894

total clf loss: 80.58676587045193 total adv loss: 56.80520083010197 749.6580483913422

evaluating...

Confusion matrix, without normalization

[[ 189 75 30]

[ 336 3334 151]

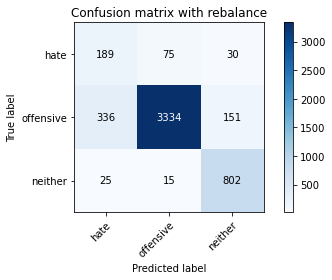
[ 25 15 802]]

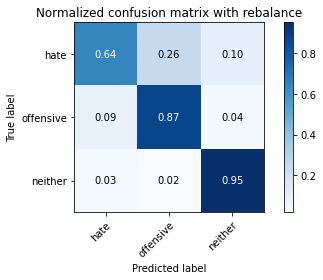
Normalized confusion matrix

[[0.64285714 0.25510204 0.10204082]

[0.0879351 0.87254645 0.03951845]

[0.02969121 0.01781473 0.95249406]]





precision recall f1-score support

hate 0.34 0.64 0.45 294

offensive 0.97 0.87 0.92 3821

neither 0.82 0.95 0.88 842

accuracy 0.87 4957

macro avg 0.71 0.82 0.75 4957

weighted avg 0.91 0.87 0.89 4957

average eval\_loss: 0.388395897893896, accuracy: 0.8725035303611055

training 3 epoch...

61.68403697013855 seconds used......

3 training finished! average train loss: -3.7820905481862743

total clf loss: 63.07502184808254 total adv loss: 48.5895384773612 758.4020440578461

evaluating...

Confusion matrix, without normalization

[[ 219 62 13]

[ 591 3131 99]

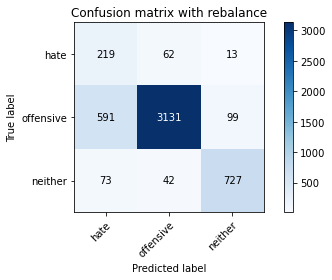
[ 73 42 727]]

Normalized confusion matrix

[[0.74489796 0.21088435 0.04421769]

[0.15467155 0.819419 0.02590945]

[0.08669834 0.04988124 0.86342043]]





precision recall f1-score support

hate 0.25 0.74 0.37 294

offensive 0.97 0.82 0.89 3821

neither 0.87 0.86 0.86 842

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: 0.4062070297077298, accuracy: 0.8224732701230583

training 4 epoch...

61.587037563323975 seconds used......

4 training finished! average train loss: -4.541936716141059

total clf loss: 50.2453752681613 total adv loss: 37.93467917852104 864.8512327671051

evaluating...

Confusion matrix, without normalization

[[ 242 32 20]

[1112 2563 146]

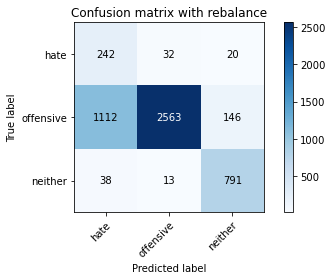
[ 38 13 791]]

Normalized confusion matrix

[[0.82312925 0.10884354 0.06802721]

[0.29102329 0.67076681 0.03820989]

[0.04513064 0.01543943 0.93942993]]





precision recall f1-score support

hate 0.17 0.82 0.29 294

offensive 0.98 0.67 0.80 3821

neither 0.83 0.94 0.88 842

accuracy 0.73 4957

macro avg 0.66 0.81 0.65 4957

weighted avg 0.91 0.73 0.78 4957

average eval\_loss: 0.6899759196225674, accuracy: 0.7254387734516845

training 5 epoch...

61.57993698120117 seconds used......

5 training finished! average train loss: -5.8614614930069235

total clf loss: 50.11031601577997 total adv loss: 33.178431146778166 1085.5986652374268

evaluating...

Confusion matrix, without normalization

[[ 224 46 24]

[ 786 2905 130]

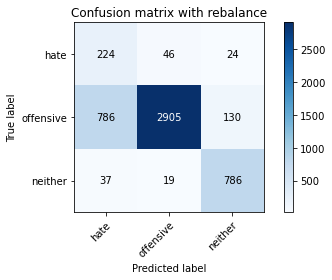
[ 37 19 786]]

Normalized confusion matrix

[[0.76190476 0.15646259 0.08163265]

[0.20570531 0.76027218 0.03402251]

[0.04394299 0.02256532 0.93349169]]





precision recall f1-score support

hate 0.21 0.76 0.33 294

offensive 0.98 0.76 0.86 3821

neither 0.84 0.93 0.88 842

accuracy 0.79 4957

macro avg 0.68 0.82 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.5560874070731863, accuracy: 0.7897922130320758

training 6 epoch...

61.625805616378784 seconds used......

6 training finished! average train loss: -6.017793856169048

total clf loss: 42.156240440905094 total adv loss: 28.228338052518666 1099.4273257255554

evaluating...

Confusion matrix, without normalization

[[ 242 37 15]

[ 959 2735 127]

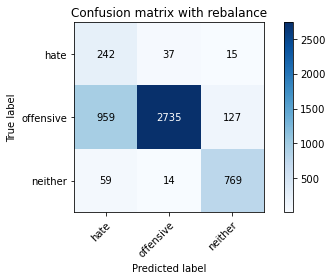
[ 59 14 769]]

Normalized confusion matrix

[[0.82312925 0.12585034 0.05102041]

[0.25098142 0.71578121 0.03323737]

[0.07007126 0.01662708 0.91330166]]





precision recall f1-score support

hate 0.19 0.82 0.31 294

offensive 0.98 0.72 0.83 3821

neither 0.84 0.91 0.88 842

accuracy 0.76 4957

macro avg 0.67 0.82 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 0.7129777917758592, accuracy: 0.7556990114988904

training 7 epoch...

61.59812879562378 seconds used......

7 training finished! average train loss: -6.263292468779269

total clf loss: 31.606384828686714 total adv loss: 17.780760211404413 1120.41015625

evaluating...

Confusion matrix, without normalization

[[ 238 43 13]

[ 882 2838 101]

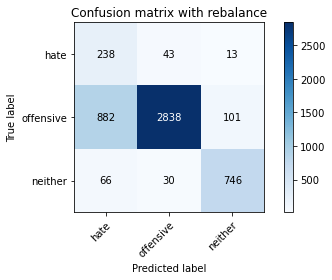
[ 66 30 746]]

Normalized confusion matrix

[[0.80952381 0.1462585 0.04421769]

[0.23082963 0.7427375 0.02643287]

[0.0783848 0.03562945 0.88598575]]





precision recall f1-score support

hate 0.20 0.81 0.32 294

offensive 0.97 0.74 0.84 3821

neither 0.87 0.89 0.88 842

accuracy 0.77 4957

macro avg 0.68 0.81 0.68 4957

weighted avg 0.91 0.77 0.82 4957

average eval\_loss: 0.6532970942977455, accuracy: 0.7710308654428082

training 8 epoch...

61.58319044113159 seconds used......

8 training finished! average train loss: -6.424383369802731

total clf loss: 24.542204838246107 total adv loss: 10.978978620260023 1134.090738773346

evaluating...

Confusion matrix, without normalization

[[ 236 43 15]

[ 858 2836 127]

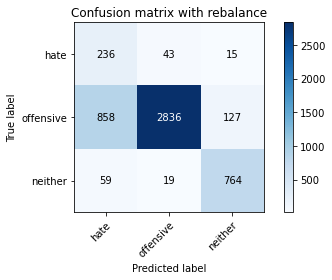
[ 59 19 764]]

Normalized confusion matrix

[[0.80272109 0.1462585 0.05102041]

[0.22454855 0.74221408 0.03323737]

[0.07007126 0.02256532 0.90736342]]





precision recall f1-score support

hate 0.20 0.80 0.33 294

offensive 0.98 0.74 0.84 3821

neither 0.84 0.91 0.87 842

accuracy 0.77 4957

macro avg 0.68 0.82 0.68 4957

weighted avg 0.91 0.77 0.82 4957

average eval\_loss: 0.617600703702098, accuracy: 0.7738551543272141

training 9 epoch...

61.61596155166626 seconds used......

9 training finished! average train loss: -6.489494323730469

total clf loss: 22.141913287341595 total adv loss: 9.338124230620451 1141.1835670471191

evaluating...

Confusion matrix, without normalization

[[ 232 47 15]

[ 837 2886 98]

[ 74 30 738]]

Normalized confusion matrix

[[0.78911565 0.15986395 0.05102041]

[0.2190526 0.75529966 0.02564774]

[0.08788599 0.03562945 0.87648456]]





precision recall f1-score support

hate 0.20 0.79 0.32 294

offensive 0.97 0.76 0.85 3821

neither 0.87 0.88 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: 0.6451522270818392, accuracy: 0.7778898527335082

training 10 epoch...

61.599390506744385 seconds used......

10 training finished! average train loss: -6.615631577564262

total clf loss: 17.880265159532428 total adv loss: 6.589456730929669 1155.74272108078

evaluating...

Confusion matrix, without normalization

[[ 230 50 14]

[ 832 2881 108]

[ 60 27 755]]

Normalized confusion matrix

[[0.78231293 0.17006803 0.04761905]

[0.21774405 0.7539911 0.02826485]

[0.07125891 0.03206651 0.89667458]]





precision recall f1-score support

hate 0.20 0.78 0.32 294

offensive 0.97 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: 0.6209444761756928, accuracy: 0.7799072019366552

4 run

training 1 epoch...

61.78586983680725 seconds used......

1 training finished! average train loss: -2.1270150496945743

total clf loss: 128.8400459587574 total adv loss: 79.11472745239735 571.6743466854095

evaluating...

Confusion matrix, without normalization

[[ 209 44 41]

[ 751 2891 179]

[ 26 12 804]]

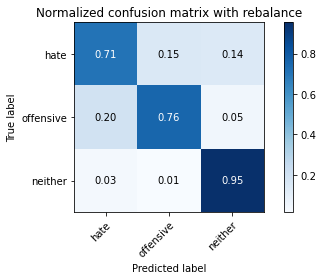
Normalized confusion matrix

[[0.71088435 0.14965986 0.13945578]

[0.19654541 0.75660822 0.04684638]

[0.03087886 0.01425178 0.95486936]]





precision recall f1-score support

hate 0.21 0.71 0.33 294

offensive 0.98 0.76 0.85 3821

neither 0.79 0.95 0.86 842

accuracy 0.79 4957

macro avg 0.66 0.81 0.68 4957

weighted avg 0.90 0.79 0.82 4957

average eval\_loss: 0.5816391055141726, accuracy: 0.7875731289086141

training 2 epoch...

61.806689500808716 seconds used......

2 training finished! average train loss: -3.5781259439144915

total clf loss: 81.84103921055794 total adv loss: 58.13573142141104 751.8363072872162

evaluating...

Confusion matrix, without normalization

[[ 210 59 25]

[ 584 3125 112]

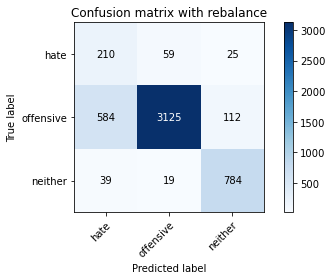
[ 39 19 784]]

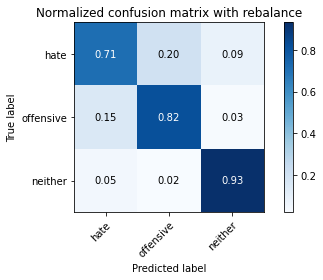
Normalized confusion matrix

[[0.71428571 0.20068027 0.08503401]

[0.15283957 0.81784873 0.0293117 ]

[0.04631829 0.02256532 0.93111639]]





precision recall f1-score support

hate 0.25 0.71 0.37 294

offensive 0.98 0.82 0.89 3821

neither 0.85 0.93 0.89 842

accuracy 0.83 4957

macro avg 0.69 0.82 0.72 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.42530526953839487, accuracy: 0.8309461367762759

training 3 epoch...

61.79312562942505 seconds used......

3 training finished! average train loss: -3.808879029681111

total clf loss: 63.53968023508787 total adv loss: 48.6052227281034 763.4632184505463

evaluating...

Confusion matrix, without normalization

[[ 204 62 28]

[ 518 3121 182]

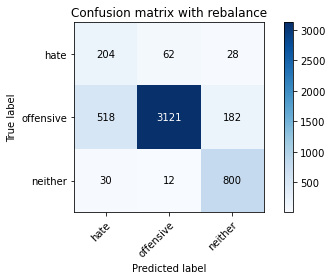
[ 30 12 800]]

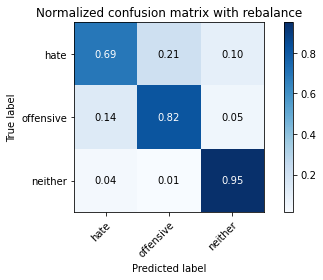
Normalized confusion matrix

[[0.69387755 0.21088435 0.0952381 ]

[0.13556661 0.81680188 0.04763151]

[0.03562945 0.01425178 0.95011876]]





precision recall f1-score support

hate 0.27 0.69 0.39 294

offensive 0.98 0.82 0.89 3821

neither 0.79 0.95 0.86 842

accuracy 0.83 4957

macro avg 0.68 0.82 0.71 4957

weighted avg 0.90 0.83 0.86 4957

average eval\_loss: 0.5182039606114549, accuracy: 0.8321565462981642

training 4 epoch...

61.79381227493286 seconds used......

4 training finished! average train loss: -5.222592767916228

total clf loss: 61.39541204273701 total adv loss: 43.349085453897715 997.8078608512878

evaluating...

Confusion matrix, without normalization

[[ 242 39 13]

[ 854 2863 104]

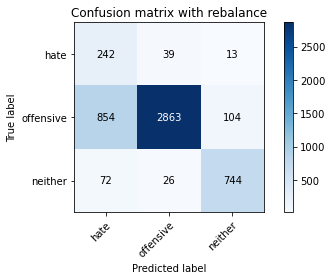
[ 72 26 744]]

Normalized confusion matrix

[[0.82312925 0.13265306 0.04421769]

[0.2235017 0.74928029 0.02721801]

[0.08551069 0.03087886 0.88361045]]





precision recall f1-score support

hate 0.21 0.82 0.33 294

offensive 0.98 0.75 0.85 3821

neither 0.86 0.88 0.87 842

accuracy 0.78 4957

macro avg 0.68 0.82 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.5640867336863472, accuracy: 0.7764777082913052

training 5 epoch...

61.82340478897095 seconds used......

5 training finished! average train loss: -5.653048071944923

total clf loss: 50.35349478572607 total adv loss: 33.24812348000705 1050.272837638855

evaluating...

Confusion matrix, without normalization

[[ 199 82 13]

[ 402 3330 89]

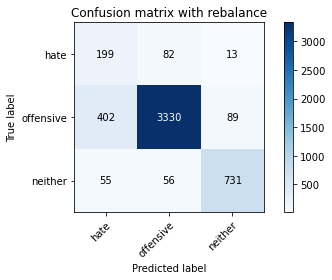
[ 55 56 731]]

Normalized confusion matrix

[[0.67687075 0.27891156 0.04421769]

[0.10520806 0.87149961 0.02329233]

[0.06532067 0.06650831 0.86817102]]





precision recall f1-score support

hate 0.30 0.68 0.42 294

offensive 0.96 0.87 0.91 3821

neither 0.88 0.87 0.87 842

accuracy 0.86 4957

macro avg 0.71 0.81 0.74 4957

weighted avg 0.91 0.86 0.88 4957

average eval\_loss: 0.3639365784493425, accuracy: 0.8593907605406496

training 6 epoch...

61.82749056816101 seconds used......

6 training finished! average train loss: -5.828950151365403

total clf loss: 42.06294792145491 total adv loss: 27.368423506617546 1066.1818461418152

evaluating...

Confusion matrix, without normalization

[[ 200 75 19]

[ 509 3192 120]

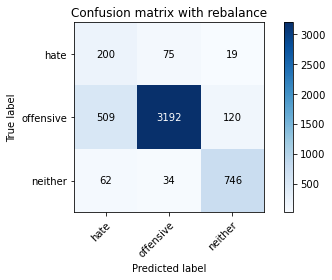
[ 62 34 746]]

Normalized confusion matrix

[[0.68027211 0.25510204 0.06462585]

[0.1332112 0.83538341 0.03140539]

[0.0736342 0.04038005 0.88598575]]





precision recall f1-score support

hate 0.26 0.68 0.38 294

offensive 0.97 0.84 0.90 3821

neither 0.84 0.89 0.86 842

accuracy 0.83 4957

macro avg 0.69 0.80 0.71 4957

weighted avg 0.90 0.83 0.86 4957

average eval\_loss: 0.44248711564127474, accuracy: 0.8347791002622554

training 7 epoch...

61.82867193222046 seconds used......

7 training finished! average train loss: -5.977961980808548

total clf loss: 33.96225295960903 total adv loss: 22.894362093880773 1079.0881133079529

evaluating...

Confusion matrix, without normalization

[[ 200 81 13]

[ 427 3296 98]

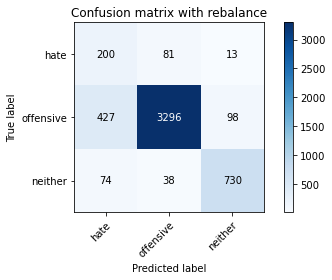
[ 74 38 730]]

Normalized confusion matrix

[[0.68027211 0.2755102 0.04421769]

[0.11175085 0.86260141 0.02564774]

[0.08788599 0.04513064 0.86698337]]





precision recall f1-score support

hate 0.29 0.68 0.40 294

offensive 0.97 0.86 0.91 3821

neither 0.87 0.87 0.87 842

accuracy 0.85 4957

macro avg 0.71 0.80 0.73 4957

weighted avg 0.91 0.85 0.87 4957

average eval\_loss: 0.40962637804688945, accuracy: 0.8525317732499496

training 8 epoch...

61.80341172218323 seconds used......

8 training finished! average train loss: -6.118185414208306

total clf loss: 27.473037637770176 total adv loss: 17.13640806544572 1090.8191509246826

evaluating...

Confusion matrix, without normalization

[[ 213 71 10]

[ 576 3178 67]

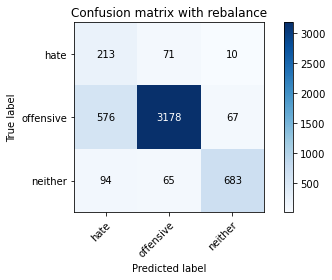
[ 94 65 683]]

Normalized confusion matrix

[[0.7244898 0.2414966 0.03401361]

[0.15074588 0.83171945 0.01753468]

[0.11163895 0.07719715 0.8111639 ]]





precision recall f1-score support

hate 0.24 0.72 0.36 294

offensive 0.96 0.83 0.89 3821

neither 0.90 0.81 0.85 842

accuracy 0.82 4957

macro avg 0.70 0.79 0.70 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.4617339856321773, accuracy: 0.8218680653621142

training 9 epoch...

61.82105493545532 seconds used......

9 training finished! average train loss: -6.175884020955939

total clf loss: 25.04375683516264 total adv loss: 13.37506907293573 1094.4949927330017

evaluating...

Confusion matrix, without normalization

[[ 212 67 15]

[ 621 3088 112]

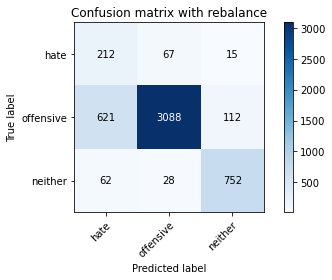
[ 62 28 752]]

Normalized confusion matrix

[[0.72108844 0.22789116 0.05102041]

[0.1625229 0.8081654 0.0293117 ]

[0.0736342 0.03325416 0.89311164]]





precision recall f1-score support

hate 0.24 0.72 0.36 294

offensive 0.97 0.81 0.88 3821

neither 0.86 0.89 0.87 842

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: 0.4963016513614885, accuracy: 0.8174298971151907

training 10 epoch...

61.71647596359253 seconds used......

10 training finished! average train loss: -6.33946050398531

total clf loss: 17.845629669725895 total adv loss: 7.068798013147898 1108.9621748924255

evaluating...

Confusion matrix, without normalization

[[ 232 52 10]

[ 733 3004 84]

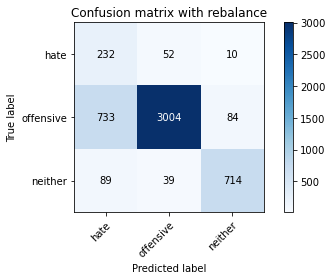
[ 89 39 714]]

Normalized confusion matrix

[[0.78911565 0.17687075 0.03401361]

[0.1918346 0.78618163 0.02198377]

[0.10570071 0.04631829 0.847981 ]]





precision recall f1-score support

hate 0.22 0.79 0.34 294

offensive 0.97 0.79 0.87 3821

neither 0.88 0.85 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: 0.5885018566773543, accuracy: 0.7968529352430905

5 run

training 1 epoch...

61.77027702331543 seconds used......

1 training finished! average train loss: -2.104686975827691

total clf loss: 122.84135618805885 total adv loss: 76.28721735626459 559.0300466418266

evaluating...

Confusion matrix, without normalization

[[ 212 48 34]

[ 732 2957 132]

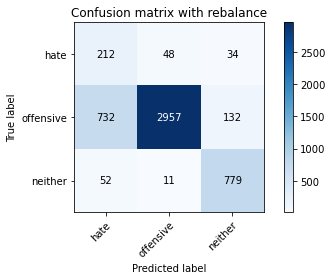
[ 52 11 779]]

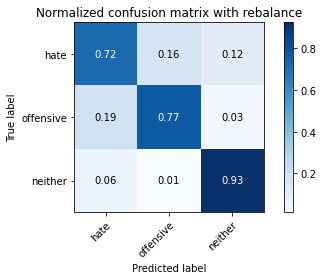
Normalized confusion matrix

[[0.72108844 0.16326531 0.11564626]

[0.19157289 0.77388118 0.03454593]

[0.06175772 0.01306413 0.92517815]]





precision recall f1-score support

hate 0.21 0.72 0.33 294

offensive 0.98 0.77 0.86 3821

neither 0.82 0.93 0.87 842

accuracy 0.80 4957

macro avg 0.67 0.81 0.69 4957

weighted avg 0.91 0.80 0.83 4957

average eval\_loss: 0.5097577272043112, accuracy: 0.7964494654024612

training 2 epoch...

61.709675312042236 seconds used......

2 training finished! average train loss: -3.419185721386246

total clf loss: 81.06585232913494 total adv loss: 56.73830324411392 722.4849144220352

evaluating...

Confusion matrix, without normalization

[[ 207 59 28]

[ 600 3067 154]

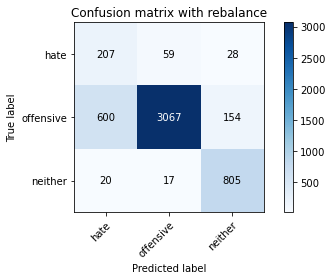
[ 20 17 805]]

Normalized confusion matrix

[[0.70408163 0.20068027 0.0952381 ]

[0.15702696 0.80266946 0.04030359]

[0.02375297 0.02019002 0.95605701]]





precision recall f1-score support

hate 0.25 0.70 0.37 294

offensive 0.98 0.80 0.88 3821

neither 0.82 0.96 0.88 842

accuracy 0.82 4957

macro avg 0.68 0.82 0.71 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.42599559836690465, accuracy: 0.8228767399636877

training 3 epoch...

61.716588497161865 seconds used......

3 training finished! average train loss: -3.6371464715366475

total clf loss: 63.79662973433733 total adv loss: 47.83165542781353 733.5803306102753

evaluating...

Confusion matrix, without normalization

[[ 218 68 8]

[ 487 3284 50]

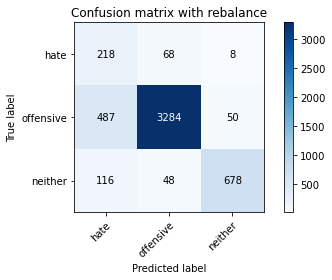
[ 116 48 678]]

Normalized confusion matrix

[[0.7414966 0.23129252 0.02721088]

[0.12745355 0.85946087 0.01308558]

[0.13776722 0.05700713 0.80522565]]





precision recall f1-score support

hate 0.27 0.74 0.39 294

offensive 0.97 0.86 0.91 3821

neither 0.92 0.81 0.86 842

accuracy 0.84 4957

macro avg 0.72 0.80 0.72 4957

weighted avg 0.92 0.84 0.87 4957

average eval\_loss: 0.4143753321540932, accuracy: 0.8432519669154731

training 4 epoch...

61.91007161140442 seconds used......

4 training finished! average train loss: -4.220332555603563

total clf loss: 50.701782904565334 total adv loss: 39.957585256546736 812.336236000061

evaluating...

Confusion matrix, without normalization

[[ 228 53 13]

[ 711 3041 69]

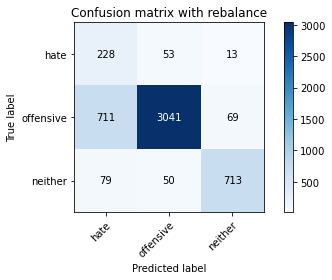
[ 79 50 713]]

Normalized confusion matrix

[[0.7755102 0.18027211 0.04421769]

[0.18607694 0.79586496 0.0180581 ]

[0.09382423 0.05938242 0.84679335]]





precision recall f1-score support

hate 0.22 0.78 0.35 294

offensive 0.97 0.80 0.87 3821

neither 0.90 0.85 0.87 842

accuracy 0.80 4957

macro avg 0.70 0.81 0.70 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.49568276425763486, accuracy: 0.8033084526931612

training 5 epoch...

61.807904958724976 seconds used......

5 training finished! average train loss: -5.651079501324927

total clf loss: 45.0281055867672 total adv loss: 29.46500823646784 1040.8277082443237

evaluating...

Confusion matrix, without normalization

[[ 214 51 29]

[ 670 2985 166]

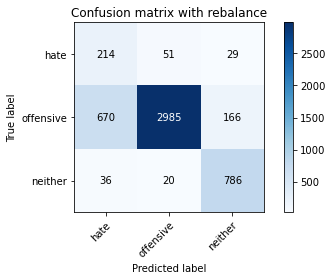
[ 36 20 786]]

Normalized confusion matrix

[[0.72789116 0.17346939 0.09863946]

[0.17534677 0.78120911 0.04344412]

[0.04275534 0.02375297 0.93349169]]





precision recall f1-score support

hate 0.23 0.73 0.35 294

offensive 0.98 0.78 0.87 3821

neither 0.80 0.93 0.86 842

accuracy 0.80 4957

macro avg 0.67 0.81 0.69 4957

weighted avg 0.90 0.80 0.84 4957

average eval\_loss: 0.5284619331359863, accuracy: 0.8039136574541053

training 6 epoch...

61.86670970916748 seconds used......

6 training finished! average train loss: -5.829747241840028

total clf loss: 36.96228337660432 total adv loss: 23.726386564783752 1057.5754494667053

evaluating...

Confusion matrix, without normalization

[[ 224 54 16]

[ 673 3052 96]

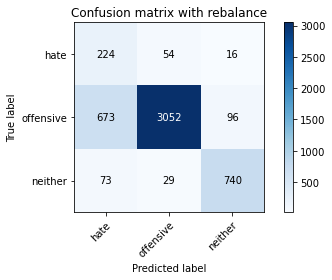
[ 73 29 740]]

Normalized confusion matrix

[[0.76190476 0.18367347 0.05442177]

[0.1761319 0.79874378 0.02512431]

[0.08669834 0.03444181 0.87885986]]





precision recall f1-score support

hate 0.23 0.76 0.35 294

offensive 0.97 0.80 0.88 3821

neither 0.87 0.88 0.87 842

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: 0.5247899759560823, accuracy: 0.8101674399838612

training 7 epoch...

61.787824392318726 seconds used......

7 training finished! average train loss: -6.071220690743965

total clf loss: 26.038885038346052 total adv loss: 14.342157293809578 1078.5597834587097

evaluating...

Confusion matrix, without normalization

[[ 215 56 23]

[ 651 3047 123]

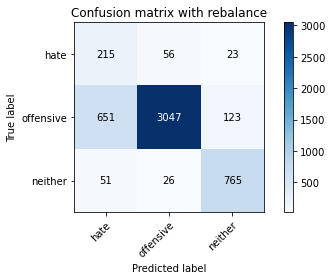
[ 51 26 765]]

Normalized confusion matrix

[[0.73129252 0.19047619 0.07823129]

[0.17037425 0.79743523 0.03219053]

[0.06057007 0.03087886 0.90855107]]





precision recall f1-score support

hate 0.23 0.73 0.36 294

offensive 0.97 0.80 0.88 3821

neither 0.84 0.91 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.91 0.81 0.85 4957

average eval\_loss: 0.5347948917037537, accuracy: 0.8123865241073229

training 8 epoch...

61.842432498931885 seconds used......

8 training finished! average train loss: -6.130343501330816

total clf loss: 22.984644684940577 total adv loss: 10.7681473151315 1082.041531085968

evaluating...

Confusion matrix, without normalization

[[ 230 43 21]

[ 897 2751 173]

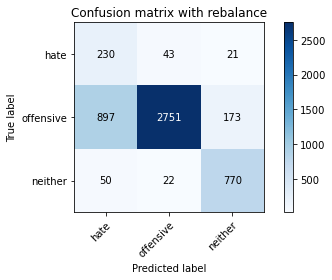
[ 50 22 770]]

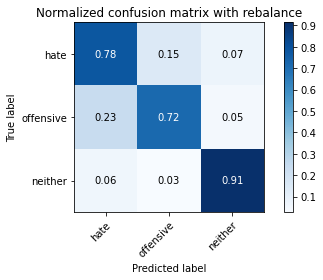
Normalized confusion matrix

[[0.78231293 0.1462585 0.07142857]

[0.2347553 0.71996859 0.04527611]

[0.05938242 0.02612827 0.91448931]]





precision recall f1-score support

hate 0.20 0.78 0.31 294

offensive 0.98 0.72 0.83 3821

neither 0.80 0.91 0.85 842

accuracy 0.76 4957

macro avg 0.66 0.81 0.66 4957

weighted avg 0.90 0.76 0.80 4957

average eval\_loss: 0.6885248759701368, accuracy: 0.756707686100464

training 9 epoch...

61.75088024139404 seconds used......

9 training finished! average train loss: -6.215172784370289

total clf loss: 20.455823495984077 total adv loss: 8.259174781502225 1091.5095462799072

evaluating...

Confusion matrix, without normalization

[[ 238 49 7]

[ 800 2934 87]

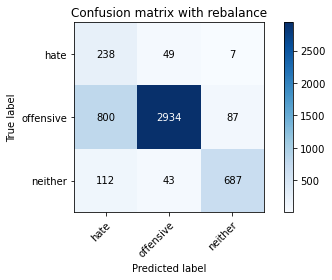
[ 112 43 687]]

Normalized confusion matrix

[[0.80952381 0.16666667 0.02380952]

[0.20936928 0.76786182 0.02276891]

[0.13301663 0.05106888 0.81591449]]





precision recall f1-score support

hate 0.21 0.81 0.33 294

offensive 0.97 0.77 0.86 3821

neither 0.88 0.82 0.85 842

accuracy 0.78 4957

macro avg 0.69 0.80 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.6117767886288704, accuracy: 0.7784950574944522

training 10 epoch...

61.59469962120056 seconds used......

10 training finished! average train loss: -6.333777937972755

total clf loss: 15.710938416421413 total adv loss: 4.714546809263993 1103.5015120506287

evaluating...

Confusion matrix, without normalization

[[ 209 69 16]

[ 580 3127 114]

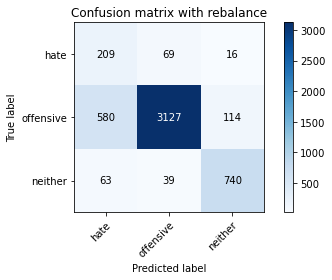
[ 63 39 740]]

Normalized confusion matrix

[[0.71088435 0.23469388 0.05442177]

[0.15179272 0.81837215 0.02983512]

[0.07482185 0.04631829 0.87885986]]





precision recall f1-score support

hate 0.25 0.71 0.36 294

offensive 0.97 0.82 0.89 3821

neither 0.85 0.88 0.86 842

accuracy 0.82 4957

macro avg 0.69 0.80 0.71 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.5105185734048006, accuracy: 0.8222715352027435