alpha = 0.05

training 1 epoch...

1543.3892152309418 seconds used......

1 training finished! average train loss: 0.2710207627933102

total clf loss: 550.3435973506421 total adv loss: 258.1884965139834

evaluating...

Confusion matrix, without normalization

[[ 27 232 35]

[ 17 3715 89]

[ 6 59 777]]

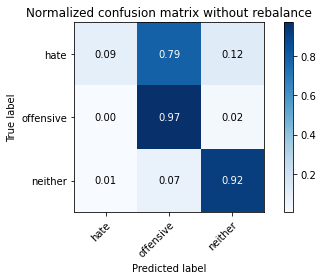
Normalized confusion matrix

[[0.09183673 0.78911565 0.11904762]

[0.0044491 0.97225857 0.02329233]

[0.00712589 0.07007126 0.92280285]]





precision recall f1-score support

hate 0.54 0.09 0.16 294

offensive 0.93 0.97 0.95 3821

neither 0.86 0.92 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.66 0.67 4957

weighted avg 0.89 0.91 0.89 4957

average eval\_loss: 0.24155023974843412, accuracy: 0.9116401049021585

training 2 epoch...

1550.3872528076172 seconds used......

2 training finished! average train loss: 0.2028776076548337

total clf loss: 410.3160943321418 total adv loss: 160.19597894722756

evaluating...

Confusion matrix, without normalization

[[ 77 205 12]

[ 52 3708 61]

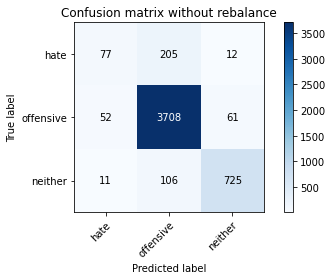
[ 11 106 725]]

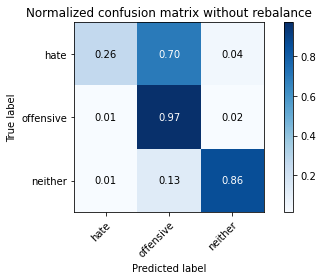
Normalized confusion matrix

[[0.26190476 0.69727891 0.04081633]

[0.013609 0.97042659 0.01596441]

[0.01306413 0.12589074 0.86104513]]





precision recall f1-score support

hate 0.55 0.26 0.35 294

offensive 0.92 0.97 0.95 3821

neither 0.91 0.86 0.88 842

accuracy 0.91 4957

macro avg 0.79 0.70 0.73 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.2526850789229596, accuracy: 0.9098244906193262

training 3 epoch...

1548.7896282672882 seconds used......

3 training finished! average train loss: 0.14360902705848588

total clf loss: 289.3258997966768 total adv loss: 90.98398100209374

evaluating...

Confusion matrix, without normalization

[[ 107 175 12]

[ 85 3676 60]

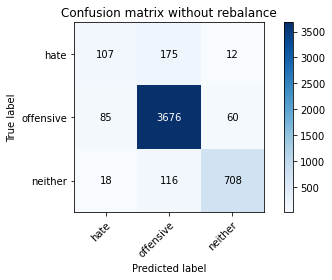
[ 18 116 708]]

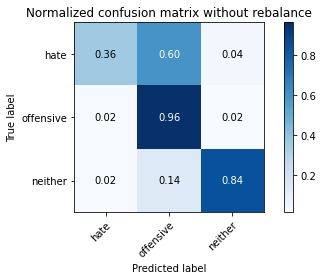
Normalized confusion matrix

[[0.36394558 0.5952381 0.04081633]

[0.02224549 0.96205182 0.0157027 ]

[0.02137767 0.13776722 0.84085511]]





precision recall f1-score support

hate 0.51 0.36 0.42 294

offensive 0.93 0.96 0.94 3821

neither 0.91 0.84 0.87 842

accuracy 0.91 4957

macro avg 0.78 0.72 0.75 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.2827138071715449, accuracy: 0.9059915271333467

training 4 epoch...

1548.7860915660858 seconds used......

4 training finished! average train loss: 0.08727696838323613

total clf loss: 175.43227023258805 total adv loss: 47.24084162181356

evaluating...

Confusion matrix, without normalization

[[ 100 177 17]

[ 96 3660 65]

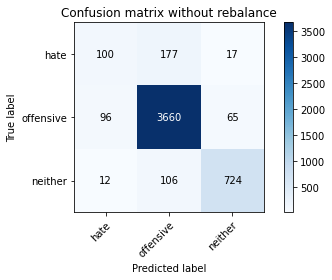
[ 12 106 724]]

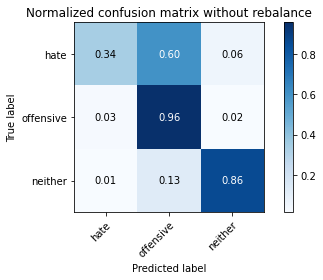
Normalized confusion matrix

[[0.34013605 0.60204082 0.05782313]

[0.02512431 0.95786443 0.01701125]

[0.01425178 0.12589074 0.85985748]]





precision recall f1-score support

hate 0.48 0.34 0.40 294

offensive 0.93 0.96 0.94 3821

neither 0.90 0.86 0.88 842

accuracy 0.90 4957

macro avg 0.77 0.72 0.74 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.3729345281728104, accuracy: 0.9045793826911438

training 5 epoch...

1548.720866203308 seconds used......

5 training finished! average train loss: 0.04618200177269677

total clf loss: 93.09540777332586 total adv loss: 30.329967281095378

evaluating...

Confusion matrix, without normalization

[[ 137 144 13]

[ 285 3470 66]

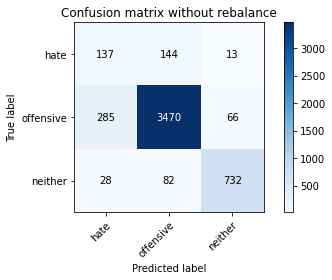
[ 28 82 732]]

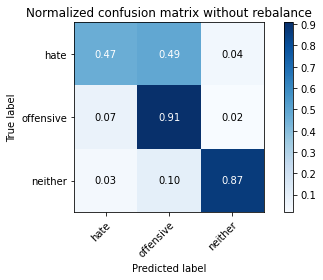
Normalized confusion matrix

[[0.46598639 0.48979592 0.04421769]

[0.0745878 0.90813923 0.01727297]

[0.03325416 0.09738717 0.86935867]]





precision recall f1-score support

hate 0.30 0.47 0.37 294

offensive 0.94 0.91 0.92 3821

neither 0.90 0.87 0.89 842

accuracy 0.88 4957

macro avg 0.72 0.75 0.73 4957

weighted avg 0.90 0.88 0.88 4957

average eval\_loss: 0.4542921523861878, accuracy: 0.8753278192455114

training 1 epoch...

894.8486814498901 seconds used......

1 training finished! average train loss: 0.27276123568591293

total clf loss: 554.0106713636778 total adv loss: 262.50281522036676

evaluating...

Confusion matrix, without normalization

[[ 34 229 31]

[ 22 3698 101]

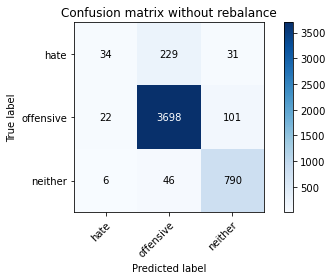
[ 6 46 790]]

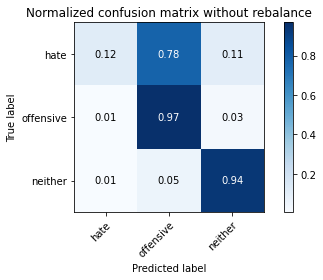
Normalized confusion matrix

[[0.11564626 0.77891156 0.10544218]

[0.00575766 0.96780947 0.02643287]

[0.00712589 0.05463183 0.93824228]]





precision recall f1-score support

hate 0.55 0.12 0.19 294

offensive 0.93 0.97 0.95 3821

neither 0.86 0.94 0.90 842

accuracy 0.91 4957

macro avg 0.78 0.67 0.68 4957

weighted avg 0.90 0.91 0.89 4957

average eval\_loss: 0.247621524290392, accuracy: 0.9122453096631027

training 2 epoch...

894.2824857234955 seconds used......

2 training finished! average train loss: 0.1991597798287087

total clf loss: 402.9312706682831 total adv loss: 159.94854775074782

evaluating...

Confusion matrix, without normalization

[[ 148 128 18]

[ 137 3601 83]

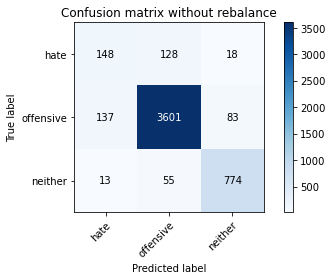
[ 13 55 774]]

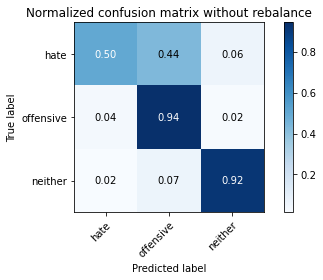
Normalized confusion matrix

[[0.50340136 0.43537415 0.06122449]

[0.03585449 0.94242345 0.02172206]

[0.01543943 0.06532067 0.9192399 ]]





precision recall f1-score support

hate 0.50 0.50 0.50 294

offensive 0.95 0.94 0.95 3821

neither 0.88 0.92 0.90 842

accuracy 0.91 4957

macro avg 0.78 0.79 0.78 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.24867789006900914, accuracy: 0.9124470445834174

training 3 epoch...

895.9516952037811 seconds used......

3 training finished! average train loss: 0.13849633773485295

total clf loss: 279.0285298292292 total adv loss: 87.8058402109549

evaluating...

Confusion matrix, without normalization

[[ 91 189 14]

[ 65 3697 59]

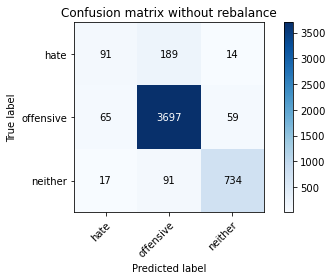
[ 17 91 734]]

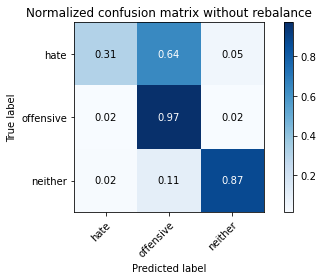
Normalized confusion matrix

[[0.30952381 0.64285714 0.04761905]

[0.01701125 0.96754776 0.01544098]

[0.02019002 0.10807601 0.87173397]]





precision recall f1-score support

hate 0.53 0.31 0.39 294

offensive 0.93 0.97 0.95 3821

neither 0.91 0.87 0.89 842

accuracy 0.91 4957

macro avg 0.79 0.72 0.74 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.2784721095468417, accuracy: 0.9122453096631027

training 4 epoch...

896.5048785209656 seconds used......

4 training finished! average train loss: 0.08129602684619727

total clf loss: 163.50534068369598 total adv loss: 45.90638960234446

evaluating...

Confusion matrix, without normalization

[[ 59 216 19]

[ 41 3706 74]

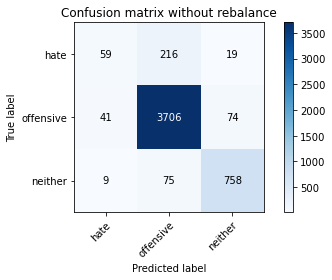
[ 9 75 758]]

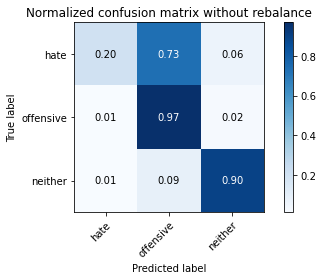
Normalized confusion matrix

[[0.20068027 0.73469388 0.06462585]

[0.01073018 0.96990317 0.01936666]

[0.01068884 0.08907363 0.90023753]]





precision recall f1-score support

hate 0.54 0.20 0.29 294

offensive 0.93 0.97 0.95 3821

neither 0.89 0.90 0.90 842

accuracy 0.91 4957

macro avg 0.79 0.69 0.71 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.3497079541864056, accuracy: 0.9124470445834174

training 5 epoch...

895.5677824020386 seconds used......

5 training finished! average train loss: 0.04053621998475524

total clf loss: 81.79063455208961 total adv loss: 28.146208254493143

evaluating...

Confusion matrix, without normalization

[[ 113 168 13]

[ 134 3634 53]

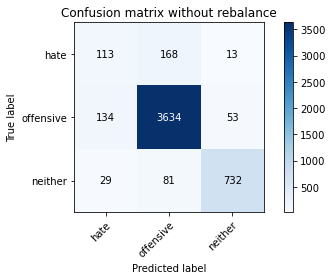
[ 29 81 732]]

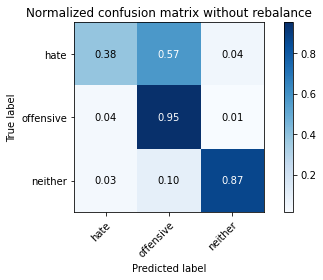
Normalized confusion matrix

[[0.38435374 0.57142857 0.04421769]

[0.03506935 0.95105993 0.01387071]

[0.03444181 0.09619952 0.86935867]]





precision recall f1-score support

hate 0.41 0.38 0.40 294

offensive 0.94 0.95 0.94 3821

neither 0.92 0.87 0.89 842

accuracy 0.90 4957

macro avg 0.75 0.73 0.74 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.4248167969936408, accuracy: 0.9035707080895703

training 6 epoch...

895.4123079776764 seconds used......

6 training finished! average train loss: 0.027321696429900434

total clf loss: 55.280105529764114 total adv loss: 22.02362820460119

evaluating...

Confusion matrix, without normalization

[[ 81 200 13]

[ 81 3693 47]

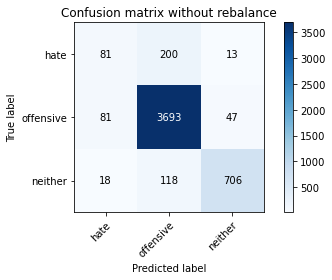
[ 18 118 706]]

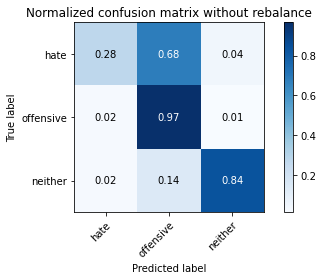
Normalized confusion matrix

[[0.2755102 0.68027211 0.04421769]

[0.02119864 0.96650092 0.01230044]

[0.02137767 0.14014252 0.83847981]]





precision recall f1-score support

hate 0.45 0.28 0.34 294

offensive 0.92 0.97 0.94 3821

neither 0.92 0.84 0.88 842

accuracy 0.90 4957

macro avg 0.76 0.69 0.72 4957

weighted avg 0.89 0.90 0.90 4957

average eval\_loss: 0.4643643392213188, accuracy: 0.903772443009885

training 7 epoch...

895.4303674697876 seconds used......

7 training finished! average train loss: 0.02042955846883333

total clf loss: 41.475092228825815 total adv loss: 19.26555809862581

evaluating...

Confusion matrix, without normalization

[[ 111 171 12]

[ 147 3631 43]

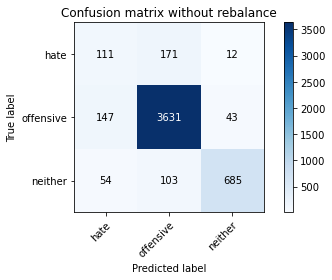
[ 54 103 685]]

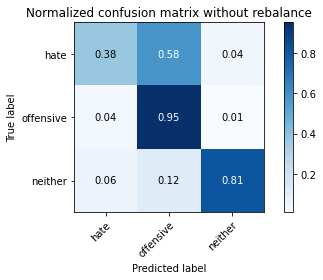
Normalized confusion matrix

[[0.37755102 0.58163265 0.04081633]

[0.0384716 0.9502748 0.0112536 ]

[0.06413302 0.12232779 0.81353919]]





precision recall f1-score support

hate 0.36 0.38 0.37 294

offensive 0.93 0.95 0.94 3821

neither 0.93 0.81 0.87 842

accuracy 0.89 4957

macro avg 0.74 0.71 0.72 4957

weighted avg 0.90 0.89 0.89 4957

average eval\_loss: 0.4974788050855251, accuracy: 0.8930804922332055

training 8 epoch...

896.0383474826813 seconds used......

8 training finished! average train loss: 0.019595005458926243

total clf loss: 39.92241336844745 total adv loss: 21.31035015462626

evaluating...

Confusion matrix, without normalization

[[ 124 157 13]

[ 153 3599 69]

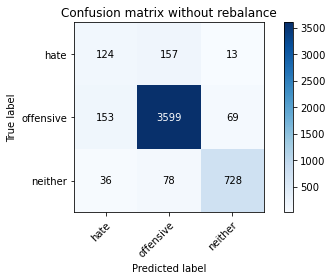
[ 36 78 728]]

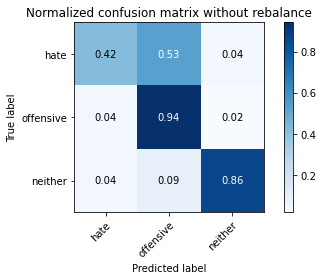
Normalized confusion matrix

[[0.42176871 0.53401361 0.04421769]

[0.04004187 0.94190003 0.0180581 ]

[0.04275534 0.09263658 0.86460808]]





precision recall f1-score support

hate 0.40 0.42 0.41 294

offensive 0.94 0.94 0.94 3821

neither 0.90 0.86 0.88 842

accuracy 0.90 4957

macro avg 0.74 0.74 0.74 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.531985000150393, accuracy: 0.8979221303207585

training 9 epoch...

895.7945051193237 seconds used......

9 training finished! average train loss: 0.017247446681388175

total clf loss: 35.09883970622468 total adv loss: 17.943061130530246

evaluating...

Confusion matrix, without normalization

[[ 84 194 16]

[ 94 3659 68]

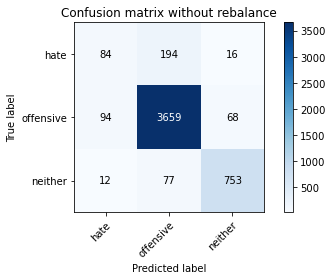
[ 12 77 753]]

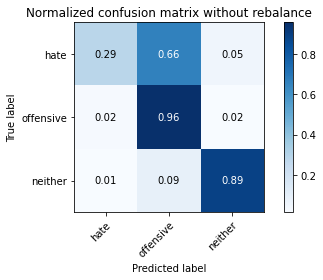
Normalized confusion matrix

[[0.28571429 0.65986395 0.05442177]

[0.02460089 0.95760272 0.01779639]

[0.01425178 0.09144893 0.89429929]]





precision recall f1-score support

hate 0.44 0.29 0.35 294

offensive 0.93 0.96 0.94 3821

neither 0.90 0.89 0.90 842

accuracy 0.91 4957

macro avg 0.76 0.71 0.73 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.5362490168037487, accuracy: 0.9070002017349204

training 10 epoch...

895.8994705677032 seconds used......

10 training finished! average train loss: 0.014419974780088546

total clf loss: 29.153904561935633 total adv loss: 11.181892354745333

evaluating...

Confusion matrix, without normalization

[[ 98 182 14]

[ 90 3659 72]

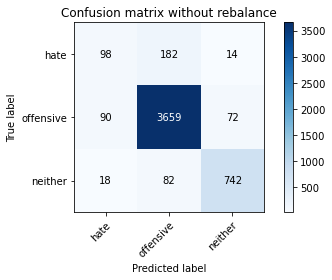
[ 18 82 742]]

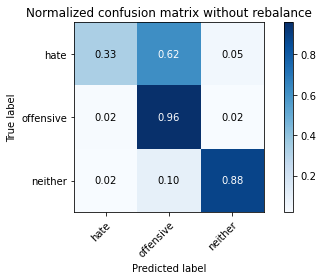
Normalized confusion matrix

[[0.33333333 0.61904762 0.04761905]

[0.02355404 0.95760272 0.01884323]

[0.02137767 0.09738717 0.88123515]]





precision recall f1-score support

hate 0.48 0.33 0.39 294

offensive 0.93 0.96 0.94 3821

neither 0.90 0.88 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.72 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.5107467338587838, accuracy: 0.9076054064958644

alpha = 0.1

unfold\_less

training 1 epoch...

894.5262899398804 seconds used......

1 training finished! average train loss: 0.269930032317668

total clf loss: 562.055046278052 total adv loss: 267.8379212634609

evaluating...

Confusion matrix, without normalization

[[ 92 176 26]

[ 90 3650 81]

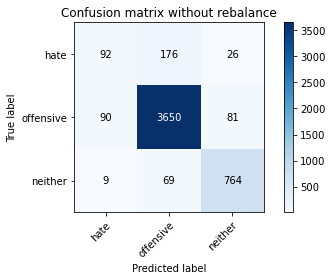
[ 9 69 764]]

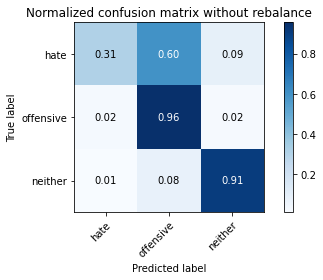
Normalized confusion matrix

[[0.31292517 0.59863946 0.08843537]

[0.02355404 0.95524732 0.02119864]

[0.01068884 0.08194774 0.90736342]]





precision recall f1-score support

hate 0.48 0.31 0.38 294

offensive 0.94 0.96 0.95 3821

neither 0.88 0.91 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.73 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.2450045044327365, accuracy: 0.9090175509380674

training 2 epoch...

893.9766256809235 seconds used......

2 training finished! average train loss: 0.20039980301719892

total clf loss: 413.7449412709102 total adv loss: 163.52131816147448

evaluating...

Confusion matrix, without normalization

[[ 96 180 18]

[ 69 3678 74]

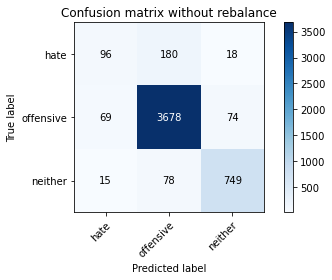
[ 15 78 749]]

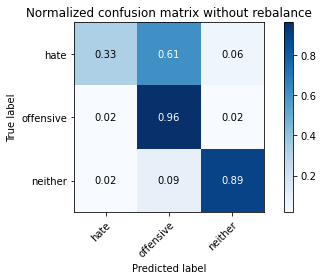
Normalized confusion matrix

[[0.32653061 0.6122449 0.06122449]

[0.0180581 0.96257524 0.01936666]

[0.01781473 0.09263658 0.88954869]]





precision recall f1-score support

hate 0.53 0.33 0.41 294

offensive 0.93 0.96 0.95 3821

neither 0.89 0.89 0.89 842

accuracy 0.91 4957

macro avg 0.79 0.73 0.75 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.2381602928334389, accuracy: 0.9124470445834174

training 3 epoch...

894.533857345581 seconds used......

3 training finished! average train loss: 0.14494213211891183

total clf loss: 296.92984641052317 total adv loss: 95.0959854088901

evaluating...

Confusion matrix, without normalization

[[ 106 165 23]

[ 78 3647 96]

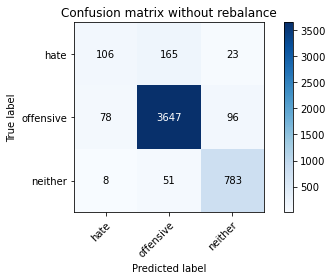
[ 8 51 783]]

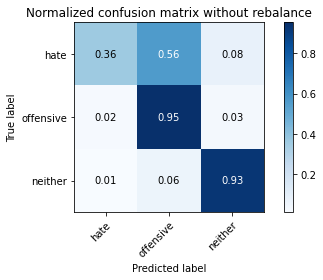
Normalized confusion matrix

[[0.36054422 0.56122449 0.07823129]

[0.0204135 0.95446218 0.02512431]

[0.00950119 0.06057007 0.92992874]]





precision recall f1-score support

hate 0.55 0.36 0.44 294

offensive 0.94 0.95 0.95 3821

neither 0.87 0.93 0.90 842

accuracy 0.92 4957

macro avg 0.79 0.75 0.76 4957

weighted avg 0.91 0.92 0.91 4957

average eval\_loss: 0.2902367986669912, accuracy: 0.9150695985475086

training 4 epoch...

893.9222071170807 seconds used......

4 training finished! average train loss: 0.08947657619780156

total clf loss: 182.29814804924536 total adv loss: 48.66097231545797

evaluating...

Confusion matrix, without normalization

[[ 100 166 28]

[ 103 3629 89]

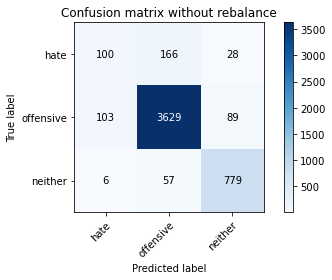
[ 6 57 779]]

Normalized confusion matrix

[[0.34013605 0.56462585 0.0952381 ]

[0.02695629 0.94975137 0.02329233]

[0.00712589 0.06769596 0.92517815]]





precision recall f1-score support

hate 0.48 0.34 0.40 294

offensive 0.94 0.95 0.95 3821

neither 0.87 0.93 0.90 842

accuracy 0.91 4957

macro avg 0.76 0.74 0.75 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.38240921005337675, accuracy: 0.9094210207786968

training 5 epoch...

893.4889395236969 seconds used......

5 training finished! average train loss: 0.04855269445064632

total clf loss: 99.53300590047002 total adv loss: 32.53012719377929

evaluating...

Confusion matrix, without normalization

[[ 119 162 13]

[ 139 3621 61]

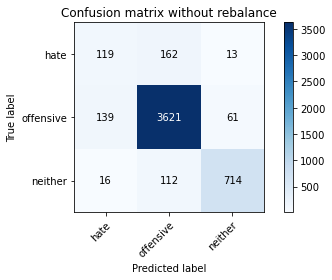
[ 16 112 714]]

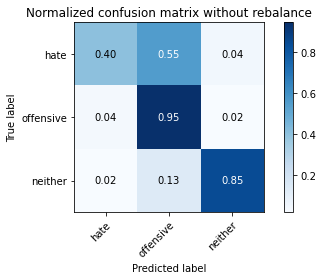
Normalized confusion matrix

[[0.4047619 0.55102041 0.04421769]

[0.03637791 0.94765768 0.01596441]

[0.01900238 0.13301663 0.847981 ]]





precision recall f1-score support

hate 0.43 0.40 0.42 294

offensive 0.93 0.95 0.94 3821

neither 0.91 0.85 0.88 842

accuracy 0.90 4957

macro avg 0.76 0.73 0.74 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.3862883630291699, accuracy: 0.8985273350817027

training 6 epoch...

894.177877664566 seconds used......

6 training finished! average train loss: 0.03072301551320327

total clf loss: 63.277076696591394 total adv loss: 23.53336897854807

evaluating...

Confusion matrix, without normalization

[[ 70 202 22]

[ 56 3683 82]

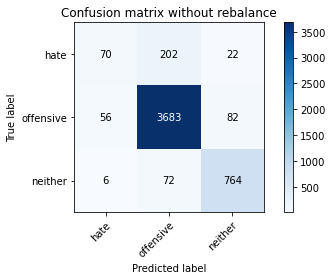
[ 6 72 764]]

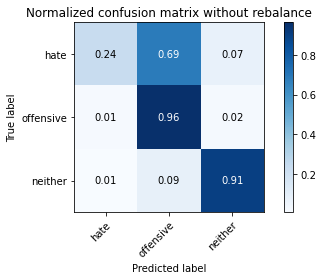
Normalized confusion matrix

[[0.23809524 0.68707483 0.07482993]

[0.01465585 0.9638838 0.02146035]

[0.00712589 0.08551069 0.90736342]]





precision recall f1-score support

hate 0.53 0.24 0.33 294

offensive 0.93 0.96 0.95 3821

neither 0.88 0.91 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.70 0.72 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.5034389885174072, accuracy: 0.9112366350615292

training 7 epoch...

895.0606684684753 seconds used......

7 training finished! average train loss: 0.023508025114839657

total clf loss: 49.12268390365716 total adv loss: 25.062699563911814

evaluating...

Confusion matrix, without normalization

[[ 120 154 20]

[ 120 3629 72]

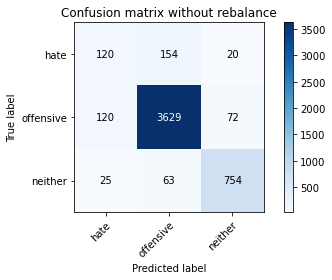
[ 25 63 754]]

Normalized confusion matrix

[[0.40816327 0.52380952 0.06802721]

[0.03140539 0.94975137 0.01884323]

[0.02969121 0.07482185 0.89548694]]





precision recall f1-score support

hate 0.45 0.41 0.43 294

offensive 0.94 0.95 0.95 3821

neither 0.89 0.90 0.89 842

accuracy 0.91 4957

macro avg 0.76 0.75 0.76 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.49089790200113265, accuracy: 0.9084123461771233

training 8 epoch...

894.8201215267181 seconds used......

8 training finished! average train loss: 0.0215716055305469

total clf loss: 45.20990133890518 total adv loss: 24.33407623502746

evaluating...

Confusion matrix, without normalization

[[ 89 186 19]

[ 96 3659 66]

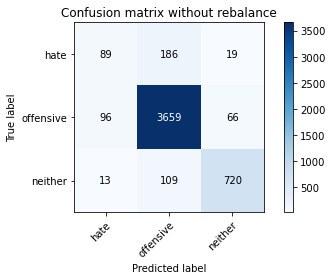
[ 13 109 720]]

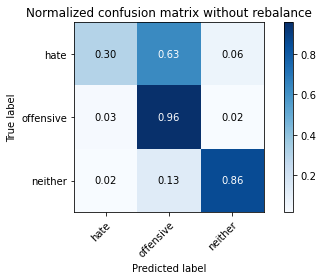
Normalized confusion matrix

[[0.30272109 0.63265306 0.06462585]

[0.02512431 0.95760272 0.01727297]

[0.01543943 0.12945368 0.85510689]]





precision recall f1-score support

hate 0.45 0.30 0.36 294

offensive 0.93 0.96 0.94 3821

neither 0.89 0.86 0.87 842

accuracy 0.90 4957

macro avg 0.76 0.71 0.73 4957

weighted avg 0.89 0.90 0.90 4957

average eval\_loss: 0.5004734331700158, accuracy: 0.9013516239661086

training 9 epoch...