alpha = 0.05

training 1 epoch...

144.29942297935486 seconds used......

1 training finished! average train loss: 0.6611683899373338

total clf loss: 116.16369591653347 total adv loss: 62.078027453273535

evaluating...

Confusion matrix, without normalization

[[ 236 39 19]

[ 872 2762 187]

[ 66 8 768]]

Normalized confusion matrix

[[0.80272109 0.13265306 0.06462585]

[0.22821251 0.72284742 0.04894007]

[0.0783848 0.00950119 0.91211401]]





precision recall f1-score support

hate 0.20 0.80 0.32 294

offensive 0.98 0.72 0.83 3821

neither 0.79 0.91 0.85 842

accuracy 0.76 4957

macro avg 0.66 0.81 0.67 4957

weighted avg 0.90 0.76 0.80 4957

average eval\_loss: 0.5669199812916978, accuracy: 0.7597337099051846

training 2 epoch...

143.7732789516449 seconds used......

2 training finished! average train loss: 0.3739197319972585

total clf loss: 65.62019690871239 total adv loss: 33.59845271334052

evaluating...

Confusion matrix, without normalization

[[ 240 39 15]

[ 963 2777 81]

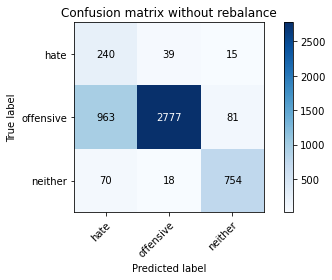
[ 70 18 754]]

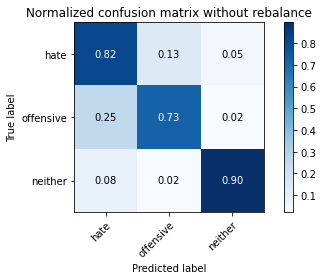
Normalized confusion matrix

[[0.81632653 0.13265306 0.05102041]

[0.25202826 0.7267731 0.02119864]

[0.08313539 0.02137767 0.89548694]]





precision recall f1-score support

hate 0.19 0.82 0.31 294

offensive 0.98 0.73 0.83 3821

neither 0.89 0.90 0.89 842

accuracy 0.76 4957

macro avg 0.69 0.81 0.68 4957

weighted avg 0.92 0.76 0.81 4957

average eval\_loss: 0.5475395995402529, accuracy: 0.7607423845067581

training 3 epoch...

143.87450408935547 seconds used......

3 training finished! average train loss: 0.2372649660911302

total clf loss: 41.842497654259205 total adv loss: 25.403769426513463

evaluating...

Confusion matrix, without normalization

[[ 219 61 14]

[ 565 3154 102]

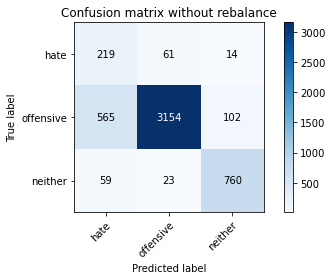
[ 59 23 760]]

Normalized confusion matrix

[[0.74489796 0.20748299 0.04761905]

[0.14786705 0.82543837 0.02669458]

[0.07007126 0.02731591 0.90261283]]





precision recall f1-score support

hate 0.26 0.74 0.39 294

offensive 0.97 0.83 0.89 3821

neither 0.87 0.90 0.88 842

accuracy 0.83 4957

macro avg 0.70 0.82 0.72 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.4663906058657073, accuracy: 0.8337704256606818

training 4 epoch...

143.89662337303162 seconds used......

4 training finished! average train loss: 0.13311698003971612

total clf loss: 23.539608562365174 total adv loss: 15.532097843126394

evaluating...

Confusion matrix, without normalization

[[ 246 33 15]

[1156 2576 89]

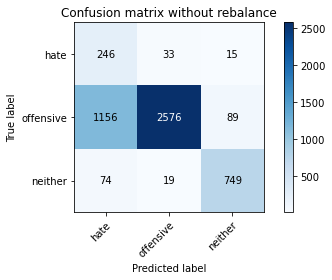
[ 74 19 749]]

Normalized confusion matrix

[[0.83673469 0.1122449 0.05102041]

[0.3025386 0.67416907 0.02329233]

[0.08788599 0.02256532 0.88954869]]





precision recall f1-score support

hate 0.17 0.84 0.28 294

offensive 0.98 0.67 0.80 3821

neither 0.88 0.89 0.88 842

accuracy 0.72 4957

macro avg 0.67 0.80 0.65 4957

weighted avg 0.91 0.72 0.78 4957

average eval\_loss: 0.7635889505186388, accuracy: 0.7203954004438168

training 5 epoch...

143.90414953231812 seconds used......

5 training finished! average train loss: 0.0661474622538237

total clf loss: 11.79329821979627 total adv loss: 9.641643193666823

evaluating...

Confusion matrix, without normalization

[[ 208 68 18]

[ 551 3185 85]

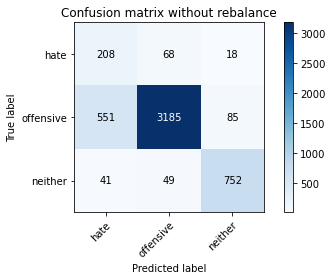
[ 41 49 752]]

Normalized confusion matrix

[[0.70748299 0.23129252 0.06122449]

[0.14420309 0.83355143 0.02224549]

[0.04869359 0.05819477 0.89311164]]





precision recall f1-score support

hate 0.26 0.71 0.38 294

offensive 0.96 0.83 0.89 3821

neither 0.88 0.89 0.89 842

accuracy 0.84 4957

macro avg 0.70 0.81 0.72 4957

weighted avg 0.91 0.84 0.86 4957

average eval\_loss: 0.6070191899507923, accuracy: 0.8361912447044584

training 1 epoch...

143.9215371608734 seconds used......

1 training finished! average train loss: 0.03493697883163671

total clf loss: 6.217400959460065 total adv loss: 4.86355108024145

evaluating...

Confusion matrix, without normalization

[[ 207 67 20]

[ 501 3223 97]

[ 39 37 766]]

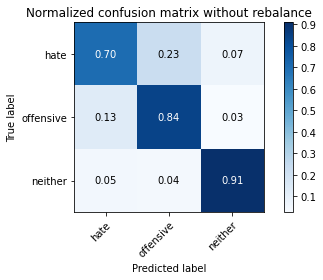
Normalized confusion matrix

[[0.70408163 0.22789116 0.06802721]

[0.13111751 0.84349647 0.02538602]

[0.04631829 0.04394299 0.90973872]]





precision recall f1-score support

hate 0.28 0.70 0.40 294

offensive 0.97 0.84 0.90 3821

neither 0.87 0.91 0.89 842

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: 0.6710841966381732, accuracy: 0.8464797256405083

training 2 epoch...

143.87230038642883 seconds used......

2 training finished! average train loss: 0.02092479581264534

total clf loss: 3.7076005202834494 total adv loss: 2.589208272700489

evaluating...

Confusion matrix, without normalization

[[ 233 41 20]

[ 972 2719 130]

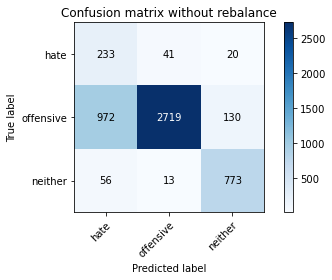
[ 56 13 773]]

Normalized confusion matrix

[[0.79251701 0.13945578 0.06802721]

[0.25438367 0.71159382 0.03402251]

[0.06650831 0.01543943 0.91805226]]





precision recall f1-score support

hate 0.18 0.79 0.30 294

offensive 0.98 0.71 0.82 3821

neither 0.84 0.92 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: 1.2672671956130332, accuracy: 0.7514625781722816

training 3 epoch...

143.85827255249023 seconds used......

3 training finished! average train loss: 0.020403418077083783

total clf loss: 3.6259969881211873 total adv loss: 2.7402500713269546

evaluating...

Confusion matrix, without normalization

[[ 230 43 21]

[ 907 2796 118]

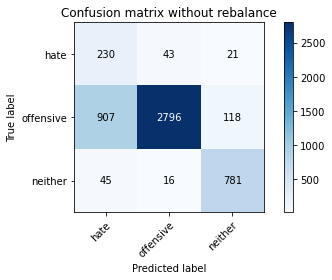
[ 45 16 781]]

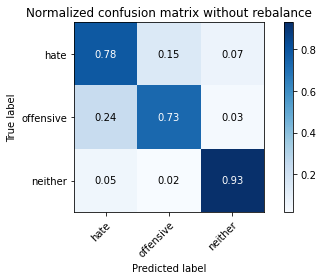
Normalized confusion matrix

[[0.78231293 0.1462585 0.07142857]

[0.23737242 0.73174562 0.03088197]

[0.05344418 0.01900238 0.92755344]]





precision recall f1-score support

hate 0.19 0.78 0.31 294

offensive 0.98 0.73 0.84 3821

neither 0.85 0.93 0.89 842

accuracy 0.77 4957

macro avg 0.67 0.81 0.68 4957

weighted avg 0.91 0.77 0.81 4957

average eval\_loss: 1.2418323810918317, accuracy: 0.7680048416380876

training 4 epoch...

143.93637561798096 seconds used......

4 training finished! average train loss: 0.016888878088000596

total clf loss: 3.0128380478709005 total adv loss: 2.496797499665263

evaluating...

Confusion matrix, without normalization

[[ 209 65 20]

[ 597 3118 106]

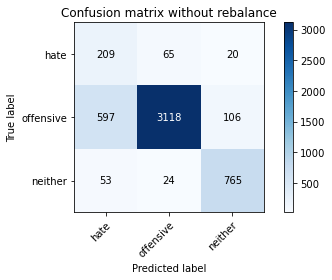
[ 53 24 765]]

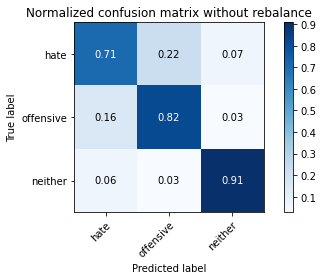
Normalized confusion matrix

[[0.71088435 0.22108844 0.06802721]

[0.15624182 0.81601675 0.02774143]

[0.06294537 0.02850356 0.90855107]]





precision recall f1-score support

hate 0.24 0.71 0.36 294

offensive 0.97 0.82 0.89 3821

neither 0.86 0.91 0.88 842

accuracy 0.83 4957

macro avg 0.69 0.81 0.71 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.9188114525224533, accuracy: 0.8254992939277789

training 5 epoch...

143.85001182556152 seconds used......

5 training finished! average train loss: 0.013160145630748882

total clf loss: 2.327820486316341 total adv loss: 1.5487117635830145

evaluating...

Confusion matrix, without normalization

[[ 197 71 26]

[ 502 3183 136]

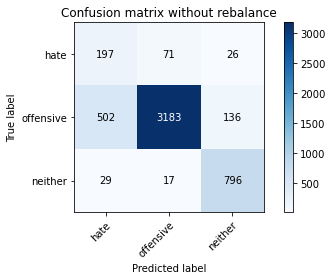
[ 29 17 796]]

Normalized confusion matrix

[[0.67006803 0.2414966 0.08843537]

[0.13137922 0.833028 0.03559278]

[0.03444181 0.02019002 0.94536817]]





precision recall f1-score support

hate 0.27 0.67 0.39 294

offensive 0.97 0.83 0.90 3821

neither 0.83 0.95 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: 0.903549212562446, accuracy: 0.8424450272342142