alpha = 0.05

training 1 epoch...

153.95786356925964 seconds used......

1 training finished! average train loss: 0.5607926122123196

total clf loss: 195.89230359345675 total adv loss: 93.24045272544026

evaluating...

Confusion matrix, without normalization

[[ 210 54 30]

[ 578 3108 135]

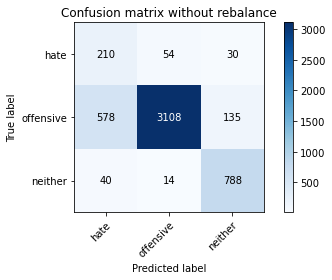
[ 40 14 788]]

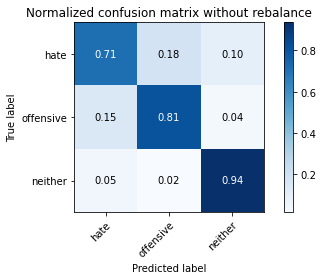
Normalized confusion matrix

[[0.71428571 0.18367347 0.10204082]

[0.1512693 0.81339963 0.03533107]

[0.04750594 0.01662708 0.93586698]]





precision recall f1-score support

hate 0.25 0.71 0.37 294

offensive 0.98 0.81 0.89 3821

neither 0.83 0.94 0.88 842

accuracy 0.83 4957

macro avg 0.69 0.82 0.71 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.5163545136475155, accuracy: 0.8283235828121848

training 2 epoch...

153.98415565490723 seconds used......

2 training finished! average train loss: 0.3659008440296671

total clf loss: 127.3083534091711 total adv loss: 50.72331490452052

evaluating...

Confusion matrix, without normalization

[[ 229 44 21]

[ 791 2900 130]

[ 34 17 791]]

Normalized confusion matrix

[[0.77891156 0.14965986 0.07142857]

[0.20701387 0.75896362 0.03402251]

[0.04038005 0.02019002 0.93942993]]





precision recall f1-score support

hate 0.22 0.78 0.34 294

offensive 0.98 0.76 0.86 3821

neither 0.84 0.94 0.89 842

accuracy 0.79 4957

macro avg 0.68 0.83 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.4845526945446768, accuracy: 0.7908008876336494

training 3 epoch...

153.90757203102112 seconds used......

3 training finished! average train loss: 0.2324993135298321

total clf loss: 80.38022507354617 total adv loss: 21.959183289523935

evaluating...

Confusion matrix, without normalization

[[ 202 67 25]

[ 512 3184 125]

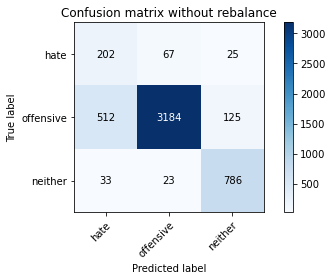
[ 33 23 786]]

Normalized confusion matrix

[[0.68707483 0.22789116 0.08503401]

[0.13399634 0.83328971 0.03271395]

[0.0391924 0.02731591 0.93349169]]





precision recall f1-score support

hate 0.27 0.69 0.39 294

offensive 0.97 0.83 0.90 3821

neither 0.84 0.93 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: 0.4009056708863711, accuracy: 0.8416380875529554

training 4 epoch...

153.72634267807007 seconds used......

4 training finished! average train loss: 0.12237890778502718

total clf loss: 42.26270236866549 total adv loss: 10.629895656409644

evaluating...

Confusion matrix, without normalization

[[ 216 60 18]

[ 701 3010 110]

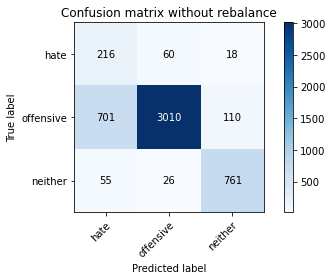
[ 55 26 761]]

Normalized confusion matrix

[[0.73469388 0.20408163 0.06122449]

[0.18345983 0.7877519 0.02878828]

[0.06532067 0.03087886 0.90380048]]





precision recall f1-score support

hate 0.22 0.73 0.34 294

offensive 0.97 0.79 0.87 3821

neither 0.86 0.90 0.88 842

accuracy 0.80 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.638383386209992, accuracy: 0.8043171272947347

training 5 epoch...

153.64642715454102 seconds used......

5 training finished! average train loss: 0.060115201086866944

total clf loss: 20.760095327859744 total adv loss: 5.216234286655435

evaluating...

Confusion matrix, without normalization

[[ 241 42 11]

[1020 2703 98]

[ 86 36 720]]

Normalized confusion matrix

[[0.81972789 0.14285714 0.03741497]

[0.26694583 0.70740644 0.02564774]

[0.10213777 0.04275534 0.85510689]]





precision recall f1-score support

hate 0.18 0.82 0.29 294

offensive 0.97 0.71 0.82 3821

neither 0.87 0.86 0.86 842

accuracy 0.74 4957

macro avg 0.67 0.79 0.66 4957

weighted avg 0.91 0.74 0.79 4957

average eval\_loss: 1.0328119865014507, accuracy: 0.7391567480330845

training 6 epoch...

153.60162329673767 seconds used......

6 training finished! average train loss: 0.0304704026999826

total clf loss: 10.533732377516571 total adv loss: 2.866499957833412

evaluating...

Confusion matrix, without normalization

[[ 206 63 25]

[ 611 3019 191]

[ 37 17 788]]

Normalized confusion matrix

[[0.70068027 0.21428571 0.08503401]

[0.15990578 0.7901073 0.04998691]

[0.04394299 0.02019002 0.93586698]]





precision recall f1-score support

hate 0.24 0.70 0.36 294

offensive 0.97 0.79 0.87 3821

neither 0.78 0.94 0.85 842

accuracy 0.81 4957

macro avg 0.67 0.81 0.70 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.87272251967601, accuracy: 0.809562235222917

training 7 epoch...

alpha = 0.09

training 1 epoch...

145.94634985923767 seconds used......

1 training finished! average train loss: 0.6427874524168103

total clf loss: 115.2385139465332 total adv loss: 59.13176829367876

evaluating...

Confusion matrix, without normalization

[[ 232 40 22]

[ 843 2868 110]

[ 59 19 764]]

Normalized confusion matrix

[[0.78911565 0.13605442 0.07482993]

[0.22062287 0.75058885 0.02878828]

[0.07007126 0.02256532 0.90736342]]





precision recall f1-score support

hate 0.20 0.79 0.32 294

offensive 0.98 0.75 0.85 3821

neither 0.85 0.91 0.88 842

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: 0.5070476346168546, accuracy: 0.7795037320960259

training 2 epoch...

145.44257760047913 seconds used......

2 training finished! average train loss: 0.3763391485339717

total clf loss: 66.60606703162193 total adv loss: 25.02302781585604

evaluating...

Confusion matrix, without normalization

[[ 234 39 21]

[ 910 2786 125]

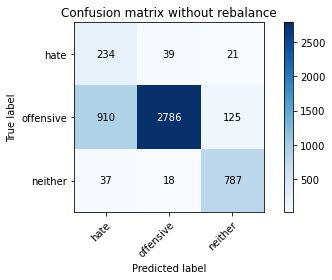
[ 37 18 787]]

Normalized confusion matrix

[[0.79591837 0.13265306 0.07142857]

[0.23815755 0.7291285 0.03271395]

[0.04394299 0.02137767 0.93467933]]





precision recall f1-score support

hate 0.20 0.80 0.32 294

offensive 0.98 0.73 0.84 3821

neither 0.84 0.93 0.89 842

accuracy 0.77 4957

macro avg 0.67 0.82 0.68 4957

weighted avg 0.91 0.77 0.81 4957

average eval\_loss: 0.5454734703075261, accuracy: 0.7680048416380876

training 1 epoch...

145.48622107505798 seconds used......

1 training finished! average train loss: 0.26623248574677966

total clf loss: 46.76528197526932 total adv loss: 13.772518613899592

evaluating...

Confusion matrix, without normalization

[[ 176 100 18]

[ 270 3451 100]

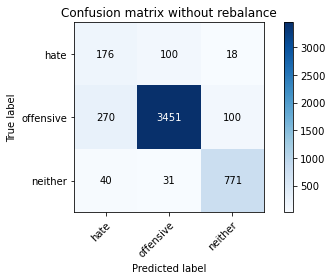
[ 40 31 771]]

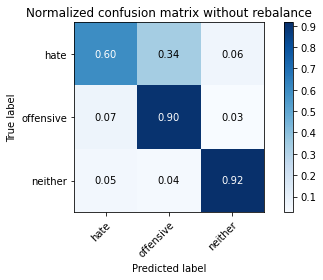
Normalized confusion matrix

[[0.59863946 0.34013605 0.06122449]

[0.07066213 0.90316671 0.02617116]

[0.04750594 0.0368171 0.91567696]]





precision recall f1-score support

hate 0.36 0.60 0.45 294

offensive 0.96 0.90 0.93 3821

neither 0.87 0.92 0.89 842

accuracy 0.89 4957

macro avg 0.73 0.81 0.76 4957

weighted avg 0.91 0.89 0.90 4957

average eval\_loss: 0.3321956364013585, accuracy: 0.8872301795440791

training 2 epoch...

145.41078758239746 seconds used......

2 training finished! average train loss: 0.15930253183414714

total clf loss: 27.785052677616477 total adv loss: 6.04799759692105

evaluating...

Confusion matrix, without normalization

[[ 236 45 13]

[ 799 2942 80]

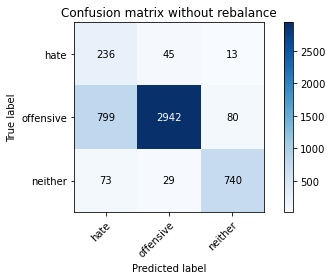
[ 73 29 740]]

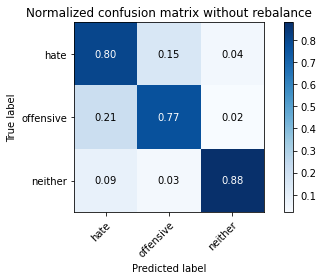
Normalized confusion matrix

[[0.80272109 0.15306122 0.04421769]

[0.20910756 0.76995551 0.02093693]

[0.08669834 0.03444181 0.87885986]]





precision recall f1-score support

hate 0.21 0.80 0.34 294

offensive 0.98 0.77 0.86 3821

neither 0.89 0.88 0.88 842

accuracy 0.79 4957

macro avg 0.69 0.82 0.69 4957

weighted avg 0.92 0.79 0.83 4957

average eval\_loss: 0.5112944245817245, accuracy: 0.79039741779302