alpha = 0.05

training 1 epoch...

835.3243033885956 seconds used......

1 training finished! average train loss: 0.27869495955265816

total clf loss: 283.6297288518399 total adv loss: 143.28657110990025

evaluating...

Confusion matrix, without normalization

[[ 109 164 21]

[ 89 3647 85]

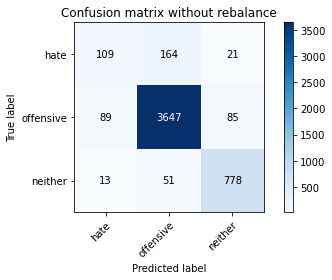
[ 13 51 778]]

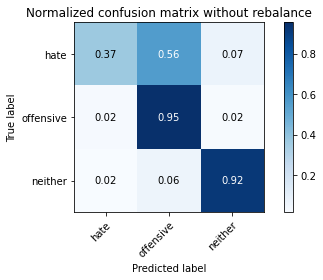
Normalized confusion matrix

[[0.3707483 0.55782313 0.07142857]

[0.02329233 0.95446218 0.02224549]

[0.01543943 0.06057007 0.9239905 ]]





precision recall f1-score support

hate 0.52 0.37 0.43 294

offensive 0.94 0.95 0.95 3821

neither 0.88 0.92 0.90 842

accuracy 0.91 4957

macro avg 0.78 0.75 0.76 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.23831696611558717, accuracy: 0.9146661287068791

training 2 epoch...

835.2520143985748 seconds used......

2 training finished! average train loss: 0.19948002486583596

total clf loss: 202.73168118484318 total adv loss: 96.94993147772038

evaluating...

Confusion matrix, without normalization

[[ 63 207 24]

[ 40 3700 81]

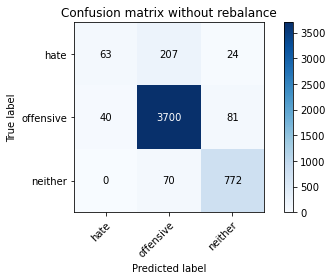
[ 0 70 772]]

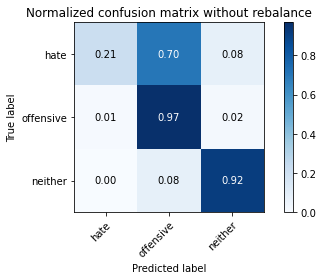
Normalized confusion matrix

[[0.21428571 0.70408163 0.08163265]

[0.01046846 0.9683329 0.02119864]

[0. 0.08313539 0.91686461]]





precision recall f1-score support

hate 0.61 0.21 0.32 294

offensive 0.93 0.97 0.95 3821

neither 0.88 0.92 0.90 842

accuracy 0.91 4957

macro avg 0.81 0.70 0.72 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.23455803083901805, accuracy: 0.9148678636271939

training 3 epoch...

835.2989115715027 seconds used......

3 training finished! average train loss: 0.1394737177235525

total clf loss: 141.69400655664504 total adv loss: 66.72157125452941

evaluating...

Confusion matrix, without normalization

[[ 132 142 20]

[ 128 3609 84]

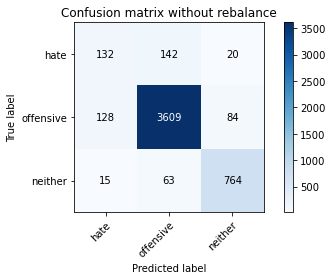
[ 15 63 764]]

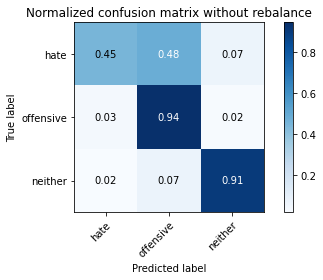
Normalized confusion matrix

[[0.44897959 0.4829932 0.06802721]

[0.03349908 0.94451714 0.02198377]

[0.01781473 0.07482185 0.90736342]]





precision recall f1-score support

hate 0.48 0.45 0.46 294

offensive 0.95 0.94 0.95 3821

neither 0.88 0.91 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.77 0.77 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.2876648415732498, accuracy: 0.9088158160177526

training 4 epoch...

835.5064167976379 seconds used......

4 training finished! average train loss: 0.08763925286503733

total clf loss: 89.33549305563793 total adv loss: 47.9470817378824

evaluating...

Confusion matrix, without normalization

[[ 112 166 16]

[ 94 3675 52]

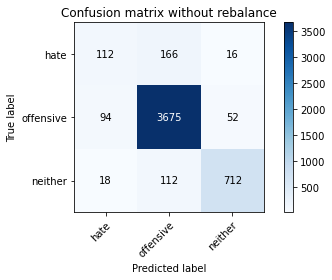
[ 18 112 712]]

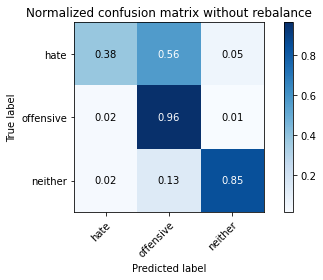
Normalized confusion matrix

[[0.38095238 0.56462585 0.05442177]

[0.02460089 0.96179011 0.013609 ]

[0.02137767 0.13301663 0.8456057 ]]





precision recall f1-score support

hate 0.50 0.38 0.43 294

offensive 0.93 0.96 0.95 3821

neither 0.91 0.85 0.88 842

accuracy 0.91 4957

macro avg 0.78 0.73 0.75 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.3130550475053174, accuracy: 0.9076054064958644

training 5 epoch...

835.4329001903534 seconds used......

5 training finished! average train loss: 0.04670001973141963

total clf loss: 47.83115908020409 total adv loss: 30.094790550505422

evaluating...

Confusion matrix, without normalization

[[ 111 161 22]

[ 128 3609 84]

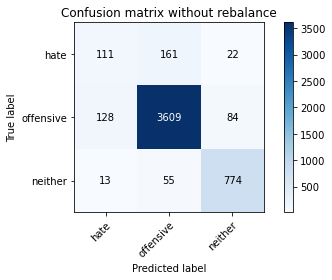
[ 13 55 774]]

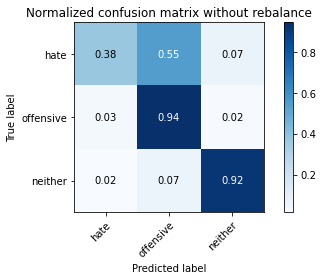
Normalized confusion matrix

[[0.37755102 0.54761905 0.07482993]

[0.03349908 0.94451714 0.02198377]

[0.01543943 0.06532067 0.9192399 ]]





precision recall f1-score support

hate 0.44 0.38 0.41 294

offensive 0.94 0.94 0.94 3821

neither 0.88 0.92 0.90 842

accuracy 0.91 4957

macro avg 0.75 0.75 0.75 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.40116044273306595, accuracy: 0.9065967318942909

training 6 epoch...

834.6913182735443 seconds used......

6 training finished! average train loss: 0.029400640390855597

total clf loss: 30.271228752739262 total adv loss: 22.115869202211798

evaluating...

Confusion matrix, without normalization

[[ 116 161 17]

[ 123 3617 81]

[ 10 67 765]]

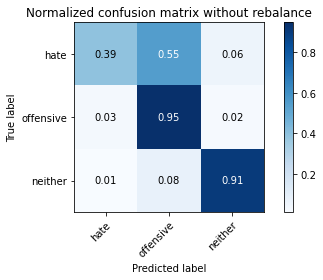
Normalized confusion matrix

[[0.39455782 0.54761905 0.05782313]

[0.03219053 0.94661083 0.02119864]

[0.01187648 0.07957245 0.90855107]]





precision recall f1-score support

hate 0.47 0.39 0.43 294

offensive 0.94 0.95 0.94 3821

neither 0.89 0.91 0.90 842

accuracy 0.91 4957

macro avg 0.76 0.75 0.76 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.4713127448772103, accuracy: 0.9074036715755497

training 7 epoch...

834.3872013092041 seconds used......

7 training finished! average train loss: 0.020434645363732875

total clf loss: 21.057769605788053 total adv loss: 15.732028082885108

evaluating...

Confusion matrix, without normalization

[[ 106 162 26]

[ 102 3639 80]

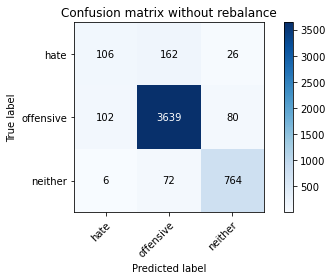
[ 6 72 764]]

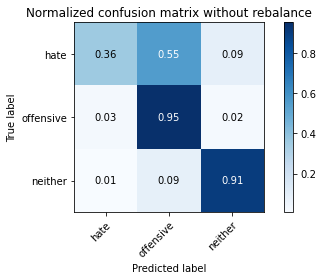
Normalized confusion matrix

[[0.36054422 0.55102041 0.08843537]

[0.02669458 0.95236849 0.02093693]

[0.00712589 0.08551069 0.90736342]]





precision recall f1-score support

hate 0.50 0.36 0.42 294

offensive 0.94 0.95 0.95 3821

neither 0.88 0.91 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.74 0.75 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.47313937304997744, accuracy: 0.9096227556990115

training 8 epoch...

835.1358470916748 seconds used......

8 training finished! average train loss: 0.017706601137543628

total clf loss: 18.31349640910048 total adv loss: 14.970960526033195

evaluating...

Confusion matrix, without normalization

[[ 90 184 20]

[ 85 3660 76]

[ 7 93 742]]

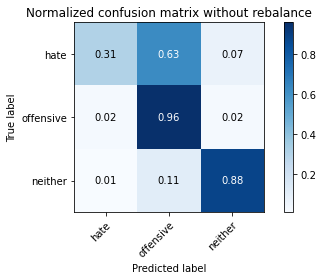
Normalized confusion matrix

[[0.30612245 0.62585034 0.06802721]

[0.02224549 0.95786443 0.01989008]

[0.00831354 0.11045131 0.88123515]]





precision recall f1-score support

hate 0.49 0.31 0.38 294

offensive 0.93 0.96 0.94 3821

neither 0.89 0.88 0.88 842

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: 0.5398669213780434, accuracy: 0.9061932620536615

training 9 epoch...

834.525045633316 seconds used......

9 training finished! average train loss: 0.017910107374665394

total clf loss: 18.51079669457249 total adv loss: 14.879403854016573

evaluating...

Confusion matrix, without normalization

[[ 154 125 15]

[ 193 3563 65]

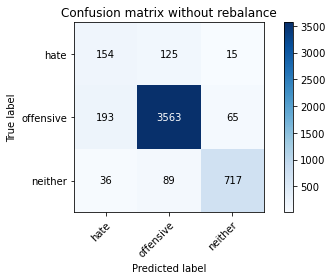
[ 36 89 717]]

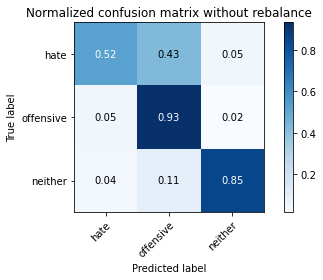
Normalized confusion matrix

[[0.52380952 0.42517007 0.05102041]

[0.05051034 0.93247841 0.01701125]

[0.04275534 0.10570071 0.85154394]]





precision recall f1-score support

hate 0.40 0.52 0.45 294

offensive 0.94 0.93 0.94 3821

neither 0.90 0.85 0.87 842

accuracy 0.89 4957

macro avg 0.75 0.77 0.76 4957

weighted avg 0.90 0.89 0.90 4957

average eval\_loss: 0.5292757738062949, accuracy: 0.8944926366754085

training 10 epoch...

834.0436918735504 seconds used......

10 training finished! average train loss: 0.013902952601986198

total clf loss: 14.361637354144477 total adv loss: 11.398166881411726

evaluating...

Confusion matrix, without normalization

[[ 117 156 21]

[ 104 3639 78]

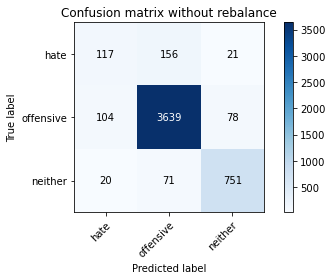
[ 20 71 751]]

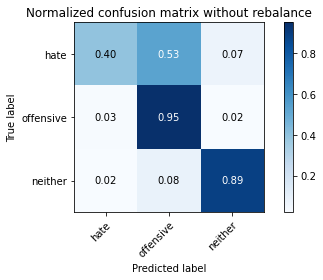
Normalized confusion matrix

[[0.39795918 0.53061224 0.07142857]

[0.02721801 0.95236849 0.0204135 ]

[0.02375297 0.08432304 0.89192399]]





precision recall f1-score support

hate 0.49 0.40 0.44 294

offensive 0.94 0.95 0.95 3821

neither 0.88 0.89 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.75 0.76 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.5499051803331573, accuracy: 0.9092192858583821