input : last 2 layer

hate as no offensive alpha=2

1 run......

Downloading: 100%

411/411 [00:00<00:00, 14.8kB/s]

Downloading: 100%

263M/263M [00:08<00:00, 31.3MB/s]

training 1 epoch...

919.8723406791687 seconds used......

1 training finished! average train loss: 0.7652887597815284

total clf loss: 289.9164544530213 total adv loss: 234.6249974835664

evaluating...

Confusion matrix, without normalization

[[ 53 214 27]

[ 43 3694 84]

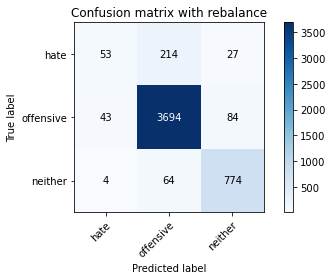
[ 4 64 774]]

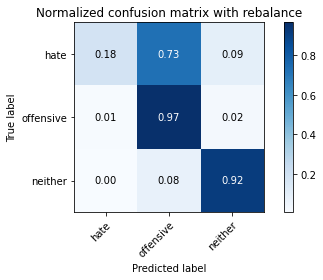
Normalized confusion matrix

[[0.18027211 0.72789116 0.09183673]

[0.0112536 0.96676263 0.02198377]

[0.00475059 0.0760095 0.9192399 ]]





precision recall f1-score support

hate 0.53 0.18 0.27 294

offensive 0.93 0.97 0.95 3821

neither 0.87 0.92 0.90 842

accuracy 0.91 4957

macro avg 0.78 0.69 0.70 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.24222079406852923, accuracy: 0.912043574742788

training 2 epoch...

920.4398205280304 seconds used......

2 training finished! average train loss: 0.593014600044567

total clf loss: 216.0346993552521 total adv loss: 186.11789162177593

evaluating...

Confusion matrix, without normalization

[[ 90 176 28]

[ 69 3665 87]

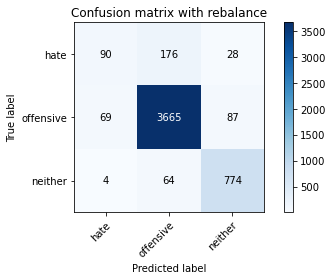
[ 4 64 774]]

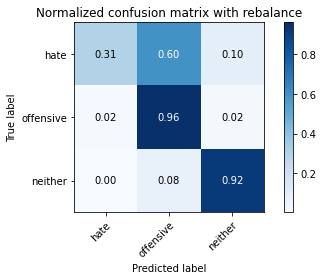
Normalized confusion matrix

[[0.30612245 0.59863946 0.0952381 ]

[0.0180581 0.95917299 0.02276891]

[0.00475059 0.0760095 0.9192399 ]]





precision recall f1-score support

hate 0.55 0.31 0.39 294

offensive 0.94 0.96 0.95 3821

neither 0.87 0.92 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: 0.2376368097555373, accuracy: 0.9136574541053056

training 3 epoch...

920.1140155792236 seconds used......

3 training finished! average train loss: 0.4795400248446141

total clf loss: 171.45280592469499 total adv loss: 152.1254494357854

evaluating...

Confusion matrix, without normalization

[[ 133 147 14]

[ 132 3622 67]

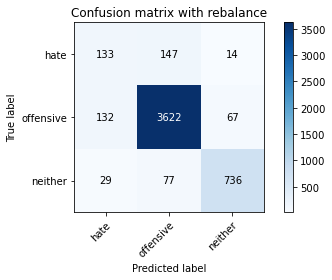
[ 29 77 736]]

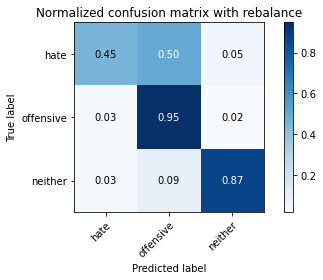
Normalized confusion matrix

[[0.45238095 0.5 0.04761905]

[0.03454593 0.94791939 0.01753468]

[0.03444181 0.09144893 0.87410926]]





precision recall f1-score support

hate 0.45 0.45 0.45 294

offensive 0.94 0.95 0.94 3821

neither 0.90 0.87 0.89 842

accuracy 0.91 4957

macro avg 0.76 0.76 0.76 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.26323321283299234, accuracy: 0.9059915271333467

training 4 epoch...

920.6521022319794 seconds used......

4 training finished! average train loss: 0.3517288428584805

total clf loss: 123.71081010159105 total adv loss: 112.60210137325339

evaluating...

Confusion matrix, without normalization

[[ 57 209 28]

[ 32 3707 82]

[ 3 71 768]]

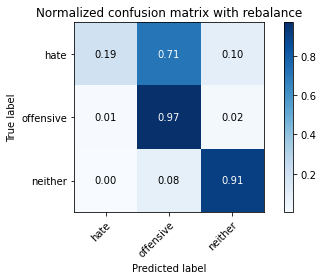
Normalized confusion matrix

[[0.19387755 0.71088435 0.0952381 ]

[0.00837477 0.97016488 0.02146035]

[0.00356295 0.08432304 0.91211401]]





precision recall f1-score support

hate 0.62 0.19 0.30 294

offensive 0.93 0.97 0.95 3821

neither 0.87 0.91 0.89 842

accuracy 0.91 4957

macro avg 0.81 0.69 0.71 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.3402369810721739, accuracy: 0.9142626588662498

training 5 epoch...

920.0021402835846 seconds used......

5 training finished! average train loss: 0.22729996661514779

total clf loss: 80.76126777881291 total adv loss: 72.36014958505984

evaluating...

Confusion matrix, without normalization

[[ 132 148 14]

[ 123 3648 50]

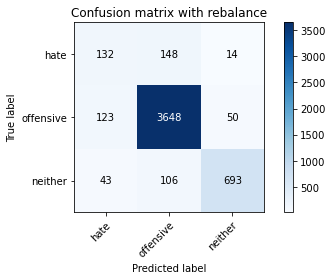
[ 43 106 693]]

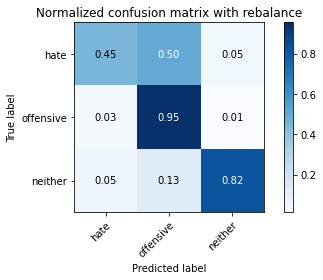
Normalized confusion matrix

[[0.44897959 0.50340136 0.04761905]

[0.03219053 0.95472389 0.01308558]

[0.05106888 0.12589074 0.82304038]]





precision recall f1-score support

hate 0.44 0.45 0.45 294

offensive 0.93 0.95 0.94 3821

neither 0.92 0.82 0.87 842

accuracy 0.90 4957

macro avg 0.76 0.74 0.75 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.33988332061750454, accuracy: 0.9023602985676821

2 run......

training 1 epoch...

919.2096168994904 seconds used......

1 training finished! average train loss: 0.7605360802676108

total clf loss: 288.8457547277212 total adv loss: 232.80301794968545

evaluating...

Confusion matrix, without normalization

[[ 101 148 45]

[ 91 3633 97]

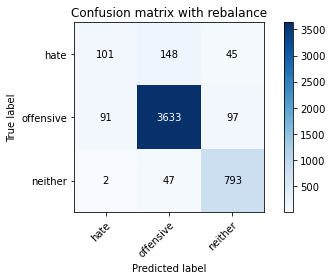
[ 2 47 793]]

Normalized confusion matrix

[[0.34353741 0.50340136 0.15306122]

[0.02381576 0.95079822 0.02538602]

[0.0023753 0.05581948 0.94180523]]





precision recall f1-score support

hate 0.52 0.34 0.41 294

offensive 0.95 0.95 0.95 3821

neither 0.85 0.94 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.75 0.75 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.24851277033467928, accuracy: 0.9132539842646762

training 2 epoch...

918.6016910076141 seconds used......

2 training finished! average train loss: 0.5893057772792095

total clf loss: 215.7202465645969 total adv loss: 184.43554210383445

evaluating...

Confusion matrix, without normalization

[[ 50 208 36]

[ 45 3677 99]

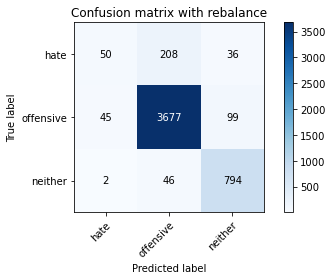
[ 2 46 794]]

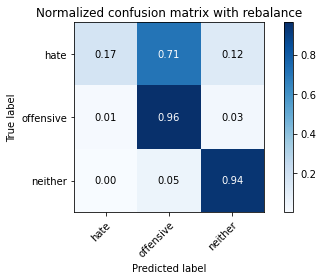
Normalized confusion matrix

[[0.17006803 0.70748299 0.12244898]

[0.01177702 0.96231353 0.02590945]

[0.0023753 0.05463183 0.94299287]]





precision recall f1-score support

hate 0.52 0.17 0.26 294

offensive 0.94 0.96 0.95 3821

neither 0.85 0.94 0.90 842

accuracy 0.91 4957

macro avg 0.77 0.69 0.70 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.24574085844740753, accuracy: 0.912043574742788

training 3 epoch...

918.5659608840942 seconds used......

3 training finished! average train loss: 0.46106499440277054

total clf loss: 166.2761057675816 total adv loss: 145.55018461449072

evaluating...

Confusion matrix, without normalization

[[ 66 203 25]

[ 45 3690 86]

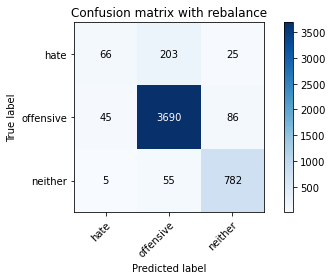
[ 5 55 782]]

Normalized confusion matrix

[[0.2244898 0.69047619 0.08503401]

[0.01177702 0.96571578 0.0225072 ]

[0.00593824 0.06532067 0.92874109]]





precision recall f1-score support

hate 0.57 0.22 0.32 294

offensive 0.93 0.97 0.95 3821

neither 0.88 0.93 0.90 842

accuracy 0.92 4957

macro avg 0.79 0.71 0.72 4957

weighted avg 0.90 0.92 0.90 4957

average eval\_loss: 0.26908747993800186, accuracy: 0.915473068388138

training 4 epoch...

919.3082141876221 seconds used......

4 training finished! average train loss: 0.3400372319836958

total clf loss: 118.79386386740953 total adv loss: 109.26153503393289

evaluating...

Confusion matrix, without normalization

[[ 92 183 19]

[ 62 3684 75]

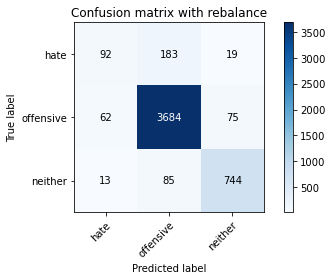
[ 13 85 744]]

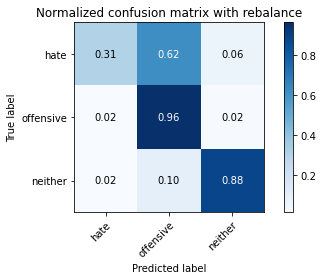
Normalized confusion matrix

[[0.31292517 0.62244898 0.06462585]

[0.01622612 0.96414551 0.01962837]

[0.01543943 0.10095012 0.88361045]]





precision recall f1-score support

hate 0.55 0.31 0.40 294

offensive 0.93 0.96 0.95 3821

neither 0.89 0.88 0.89 842

accuracy 0.91 4957

macro avg 0.79 0.72 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.30171726553160094, accuracy: 0.9118418398224732

training 5 epoch...

920.7302219867706 seconds used......

5 training finished! average train loss: 0.2175026904984524

total clf loss: 76.07710188592318 total adv loss: 69.84278371139953

evaluating...

Confusion matrix, without normalization

[[ 74 198 22]

[ 51 3705 65]

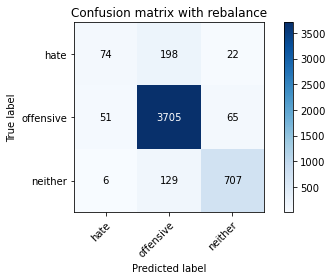
[ 6 129 707]]

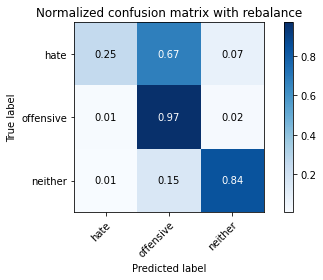
Normalized confusion matrix

[[0.25170068 0.67346939 0.07482993]

[0.01334729 0.96964146 0.01701125]

[0.00712589 0.15320665 0.83966746]]





precision recall f1-score support

hate 0.56 0.25 0.35 294

offensive 0.92 0.97 0.94 3821

neither 0.89 0.84 0.86 842

accuracy 0.90 4957

macro avg 0.79 0.69 0.72 4957

weighted avg 0.89 0.90 0.89 4957

average eval\_loss: 0.385485984582638, accuracy: 0.9049828525317732

3 run......

training 1 epoch...

input : last layer

hate as no offensive alpha=2

1 run......

Downloading: 100%

411/411 [00:00<00:00, 14.9kB/s]

Downloading: 100%

263M/263M [00:12<00:00, 23.0MB/s]

training 1 epoch...

917.8116798400879 seconds used......

1 training finished! average train loss: 0.7543130769514509

total clf loss: 291.51428002491593 total adv loss: 228.38214567489922

evaluating...

Confusion matrix, without normalization

[[ 58 197 39]

[ 61 3663 97]

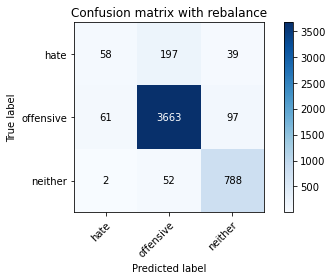
[ 2 52 788]]

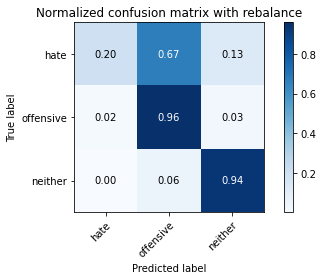
Normalized confusion matrix

[[0.19727891 0.67006803 0.13265306]

[0.01596441 0.95864957 0.02538602]

[0.0023753 0.06175772 0.93586698]]





precision recall f1-score support

hate 0.48 0.20 0.28 294

offensive 0.94 0.96 0.95 3821

neither 0.85 0.94 0.89 842

accuracy 0.91 4957

macro avg 0.76 0.70 0.71 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.2447829834606138, accuracy: 0.9096227556990115

training 2 epoch...

919.048647403717 seconds used......

2 training finished! average train loss: 0.5847107675940459

total clf loss: 213.58429467119277 total adv loss: 183.22439345531166

evaluating...

Confusion matrix, without normalization

[[ 117 158 19]

[ 94 3650 77]

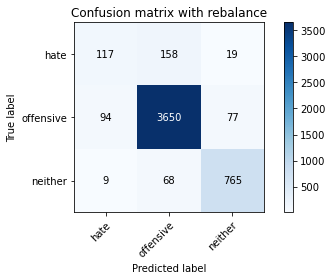
[ 9 68 765]]

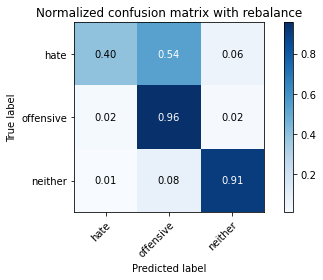
Normalized confusion matrix

[[0.39795918 0.53741497 0.06462585]

[0.02460089 0.95524732 0.02015179]

[0.01068884 0.0807601 0.90855107]]





precision recall f1-score support

hate 0.53 0.40 0.46 294

offensive 0.94 0.96 0.95 3821

neither 0.89 0.91 0.90 842

accuracy 0.91 4957

macro avg 0.79 0.75 0.77 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.24450721338601603, accuracy: 0.9142626588662498

training 3 epoch...

919.3180072307587 seconds used......

3 training finished! average train loss: 0.45837839709917805

total clf loss: 165.46655117720366 total adv loss: 144.6224093255587

evaluating...

Confusion matrix, without normalization

[[ 81 185 28]

[ 53 3658 110]

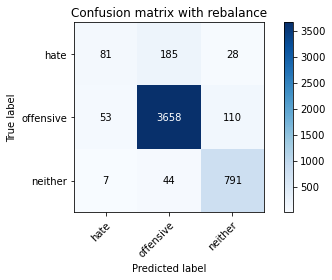
[ 7 44 791]]

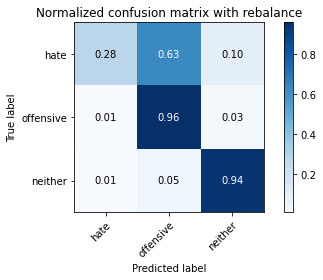
Normalized confusion matrix

[[0.2755102 0.6292517 0.0952381 ]

[0.01387071 0.95734101 0.02878828]

[0.00831354 0.05225653 0.93942993]]





precision recall f1-score support

hate 0.57 0.28 0.37 294

offensive 0.94 0.96 0.95 3821

neither 0.85 0.94 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: 0.28843632672590414, accuracy: 0.9138591890256204

training 4 epoch...

918.9000897407532 seconds used......

4 training finished! average train loss: 0.34393490279786243

total clf loss: 122.76639784220606 total adv loss: 109.20851262961514

evaluating...

Confusion matrix, without normalization

[[ 83 197 14]

[ 45 3707 69]

[ 15 96 731]]

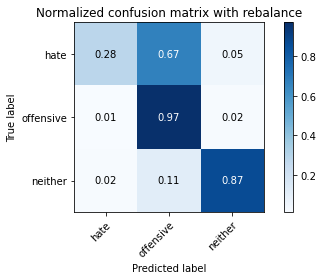
Normalized confusion matrix

[[0.28231293 0.67006803 0.04761905]

[0.01177702 0.97016488 0.0180581 ]

[0.01781473 0.11401425 0.86817102]]





precision recall f1-score support

hate 0.58 0.28 0.38 294

offensive 0.93 0.97 0.95 3821

neither 0.90 0.87 0.88 842

accuracy 0.91 4957

macro avg 0.80 0.71 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.28963744831309024, accuracy: 0.912043574742788

training 5 epoch...

919.1616687774658 seconds used......

5 training finished! average train loss: 0.2207335118434839

total clf loss: 79.34377599682193 total adv loss: 69.81193403567886

evaluating...

Confusion matrix, without normalization

[[ 91 187 16]

[ 74 3688 59]

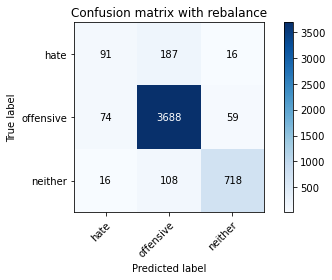
[ 16 108 718]]

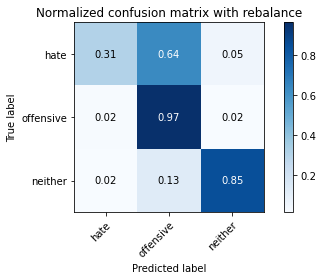
Normalized confusion matrix

[[0.30952381 0.63605442 0.05442177]

[0.01936666 0.96519236 0.01544098]

[0.01900238 0.12826603 0.85273159]]





precision recall f1-score support

hate 0.50 0.31 0.38 294

offensive 0.93 0.97 0.95 3821

neither 0.91 0.85 0.88 842

accuracy 0.91 4957

macro avg 0.78 0.71 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.33277468813110084, accuracy: 0.9072019366552351

2 run......

training 1 epoch...