alpha = 0.05

1 run......

training 1 epoch...

837.0293891429901 seconds used......

1 training finished! average train loss: 0.28464657148428923

total clf loss: 289.31321132928133 total adv loss: 138.87625136529095

evaluating...

Confusion matrix, without normalization

[[ 75 180 39]

[ 77 3636 108]

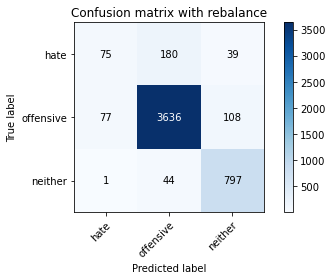
[ 1 44 797]]

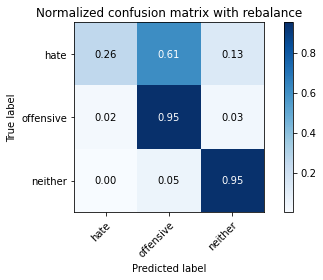
Normalized confusion matrix

[[0.25510204 0.6122449 0.13265306]

[0.02015179 0.95158336 0.02826485]

[0.00118765 0.05225653 0.94655582]]





precision recall f1-score support

hate 0.49 0.26 0.34 294

offensive 0.94 0.95 0.95 3821

neither 0.84 0.95 0.89 842

accuracy 0.91 4957

macro avg 0.76 0.72 0.72 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.2419754937366252, accuracy: 0.9094210207786968

training 2 epoch...

837.2523546218872 seconds used......

2 training finished! average train loss: 0.1999103465162578

total clf loss: 202.420177009888 total adv loss: 82.18226120075997

evaluating...

Confusion matrix, without normalization

[[ 128 146 20]

[ 97 3643 81]

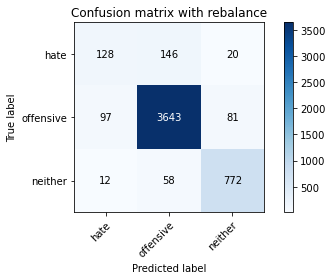
[ 12 58 772]]

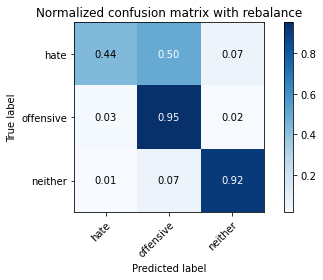
Normalized confusion matrix

[[0.43537415 0.49659864 0.06802721]

[0.02538602 0.95341534 0.02119864]

[0.01425178 0.06888361 0.91686461]]





precision recall f1-score support

hate 0.54 0.44 0.48 294

offensive 0.95 0.95 0.95 3821

neither 0.88 0.92 0.90 842

accuracy 0.92 4957

macro avg 0.79 0.77 0.78 4957

weighted avg 0.91 0.92 0.91 4957

average eval\_loss: 0.2353556936065997, accuracy: 0.9164817429897115

training 3 epoch...

836.9477479457855 seconds used......

3 training finished! average train loss: 0.14369008196574154

total clf loss: 144.74423090089113 total adv loss: 44.07338380429428

evaluating...

Confusion matrix, without normalization

[[ 95 178 21]

[ 81 3642 98]

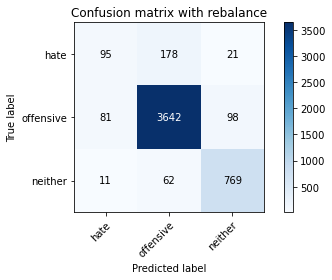
[ 11 62 769]]

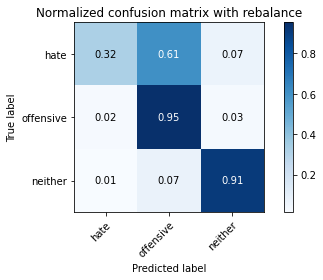
Normalized confusion matrix

[[0.32312925 0.60544218 0.07142857]

[0.02119864 0.95315362 0.02564774]

[0.01306413 0.0736342 0.91330166]]





precision recall f1-score support

hate 0.51 0.32 0.40 294

offensive 0.94 0.95 0.95 3821

neither 0.87 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: 0.2720976469046887, accuracy: 0.9090175509380674

training 4 epoch...

837.2324137687683 seconds used......

4 training finished! average train loss: 0.09011409159105707

total clf loss: 90.53108045132831 total adv loss: 22.75803059479449

evaluating...

Confusion matrix, without normalization

[[ 128 148 18]

[ 132 3612 77]

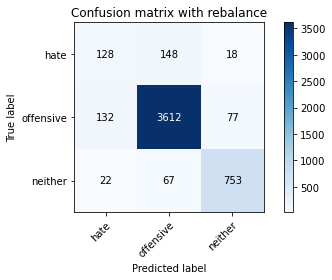
[ 22 67 753]]

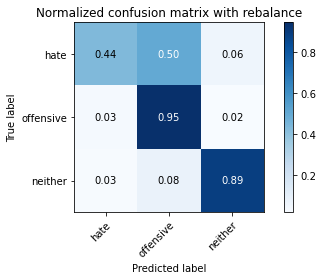
Normalized confusion matrix

[[0.43537415 0.50340136 0.06122449]

[0.03454593 0.94530228 0.02015179]

[0.02612827 0.07957245 0.89429929]]





precision recall f1-score support

hate 0.45 0.44 0.44 294

offensive 0.94 0.95 0.94 3821

neither 0.89 0.89 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: 0.32962785543866396, accuracy: 0.9063949969739762

training 5 epoch...

837.5793323516846 seconds used......

5 training finished! average train loss: 0.04999637792535659

total clf loss: 50.371369727974525 total adv loss: 15.499256886880175

evaluating...

Confusion matrix, without normalization

[[ 89 186 19]

[ 83 3654 84]

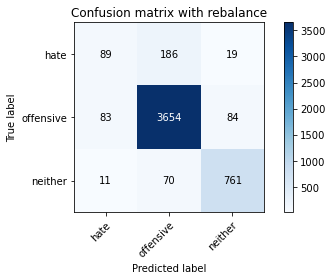
[ 11 70 761]]

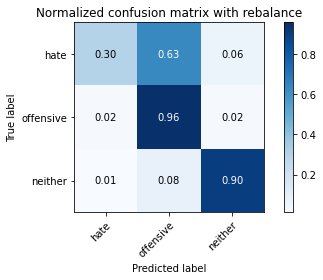
Normalized confusion matrix

[[0.30272109 0.63265306 0.06462585]

[0.02172206 0.95629416 0.02198377]

[0.01306413 0.08313539 0.90380048]]





precision recall f1-score support

hate 0.49 0.30 0.37 294

offensive 0.93 0.96 0.95 3821

neither 0.88 0.90 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: 0.38312444945674334, accuracy: 0.908614081097438

training 6 epoch...

837.516242980957 seconds used......

6 training finished! average train loss: 0.030518306240200247

total clf loss: 30.738646004872862 total adv loss: 9.28972498331433

evaluating...

Confusion matrix, without normalization

[[ 86 189 19]

[ 65 3671 85]

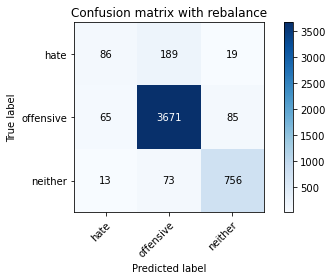
[ 13 73 756]]

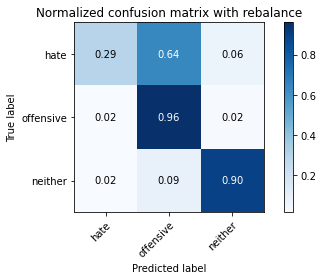
Normalized confusion matrix

[[0.29251701 0.64285714 0.06462585]

[0.01701125 0.96074326 0.02224549]

[0.01543943 0.08669834 0.89786223]]





precision recall f1-score support

hate 0.52 0.29 0.38 294

offensive 0.93 0.96 0.95 3821

neither 0.88 0.90 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.72 0.74 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.4596008576069336, accuracy: 0.9104296953802703

training 7 epoch...

838.1661374568939 seconds used......

7 training finished! average train loss: 0.020350294944503935

total clf loss: 20.671705733308045 total adv loss: 9.684263327052678

evaluating...

Confusion matrix, without normalization

[[ 114 155 25]

[ 116 3618 87]

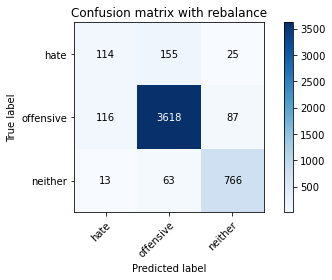
[ 13 63 766]]

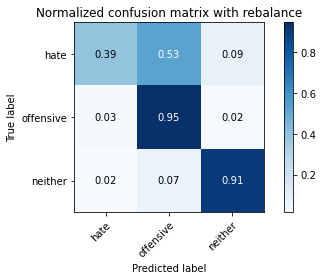
Normalized confusion matrix

[[0.3877551 0.52721088 0.08503401]

[0.03035854 0.94687255 0.02276891]

[0.01543943 0.07482185 0.90973872]]





precision recall f1-score support

hate 0.47 0.39 0.42 294

offensive 0.94 0.95 0.95 3821

neither 0.87 0.91 0.89 842

accuracy 0.91 4957

macro avg 0.76 0.75 0.75 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.5071451773370251, accuracy: 0.9074036715755497

training 8 epoch...

838.0009202957153 seconds used......

8 training finished! average train loss: 0.0162963646078116

total clf loss: 16.513961455537356 total adv loss: 6.959356515162092

evaluating...

Confusion matrix, without normalization

[[ 114 165 15]

[ 117 3655 49]

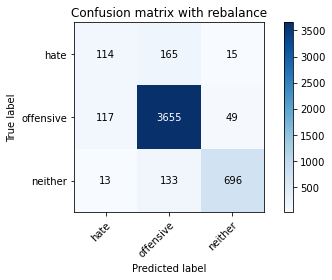
[ 13 133 696]]

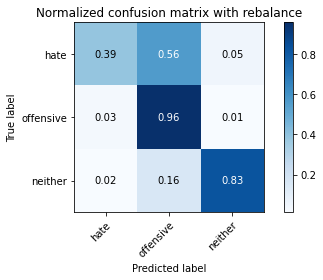
Normalized confusion matrix

[[0.3877551 0.56122449 0.05102041]

[0.03062026 0.95655588 0.01282387]

[0.01543943 0.15795724 0.82660333]]





precision recall f1-score support

hate 0.47 0.39 0.42 294

offensive 0.92 0.96 0.94 3821

neither 0.92 0.83 0.87 842

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: 0.514416347436582, accuracy: 0.9007464192051644

training 9 epoch...

837.7613418102264 seconds used......

9 training finished! average train loss: 0.01440985708891166

total clf loss: 14.617996088341897 total adv loss: 6.468358395782333

evaluating...

Confusion matrix, without normalization

[[ 123 158 13]

[ 153 3611 57]

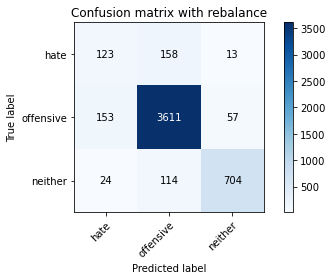
[ 24 114 704]]

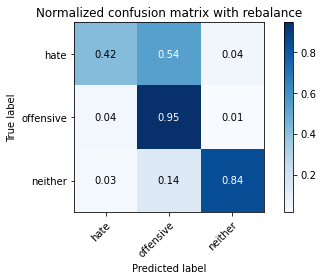
Normalized confusion matrix

[[0.41836735 0.53741497 0.04421769]

[0.04004187 0.94504057 0.01491756]

[0.02850356 0.13539192 0.83610451]]





precision recall f1-score support

hate 0.41 0.42 0.41 294

offensive 0.93 0.95 0.94 3821

neither 0.91 0.84 0.87 842

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: 0.5263882433471794, accuracy: 0.8952995763566673

training 10 epoch...

837.1599805355072 seconds used......

10 training finished! average train loss: 0.017124508082177622

total clf loss: 17.467230169655522 total adv loss: 9.594363185302235

evaluating...

Confusion matrix, without normalization

[[ 114 158 22]

[ 123 3595 103]

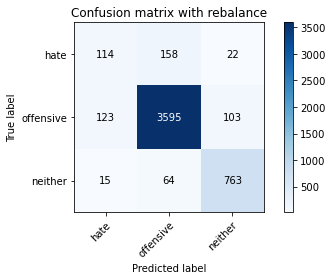
[ 15 64 763]]

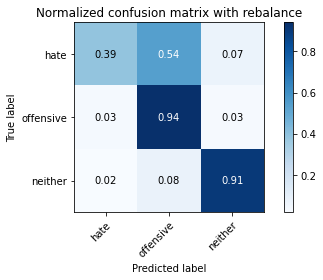
Normalized confusion matrix

[[0.3877551 0.53741497 0.07482993]

[0.03219053 0.94085318 0.02695629]

[0.01781473 0.0760095 0.90617577]]





precision recall f1-score support

hate 0.45 0.39 0.42 294

offensive 0.94 0.94 0.94 3821

neither 0.86 0.91 0.88 842

accuracy 0.90 4957

macro avg 0.75 0.74 0.75 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.5244781042301838, accuracy: 0.9021585636473674

another try

1 run......

Downloading: 100%

411/411 [00:00<00:00, 14.7kB/s]

Downloading: 100%

263M/263M [00:12<00:00, 22.4MB/s]

training 1 epoch...

842.9603390693665 seconds used......

1 training finished! average train loss: 0.2789384000559127

total clf loss: 283.54052466899157 total adv loss: 136.67262809211388

evaluating...

Confusion matrix, without normalization

[[ 49 219 26]

[ 43 3675 103]

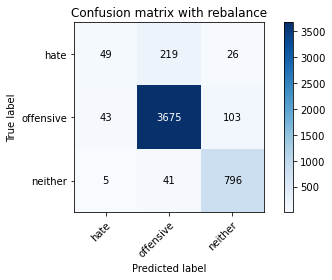
[ 5 41 796]]

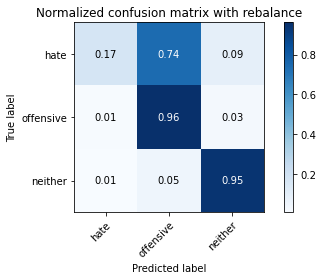
Normalized confusion matrix

[[0.16666667 0.74489796 0.08843537]

[0.0112536 0.96179011 0.02695629]

[0.00593824 0.04869359 0.94536817]]





precision recall f1-score support

hate 0.51 0.17 0.25 294

offensive 0.93 0.96 0.95 3821

neither 0.86 0.95 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: 0.2365874847427251, accuracy: 0.9118418398224732

training 2 epoch...

842.8870191574097 seconds used......

2 training finished! average train loss: 0.19997261596783514

total clf loss: 202.3963672062382 total adv loss: 80.47064572812815

evaluating...

Confusion matrix, without normalization

[[ 141 131 22]

[ 135 3599 87]

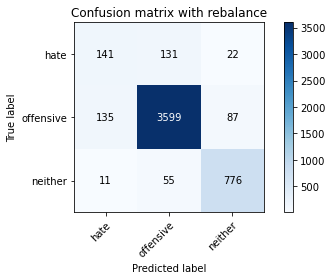
[ 11 55 776]]

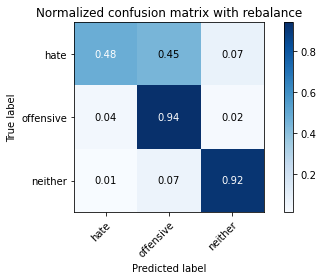
Normalized confusion matrix

[[0.47959184 0.44557823 0.07482993]

[0.03533107 0.94190003 0.02276891]

[0.01306413 0.06532067 0.9216152 ]]





precision recall f1-score support

hate 0.49 0.48 0.49 294

offensive 0.95 0.94 0.95 3821

neither 0.88 0.92 0.90 842

accuracy 0.91 4957

macro avg 0.77 0.78 0.78 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.2507017897441983, accuracy: 0.9110349001412145

training 3 epoch...

843.5022361278534 seconds used......

3 training finished! average train loss: 0.14638083775747462

total clf loss: 147.5840409793891 total adv loss: 47.48500284855618

evaluating...

Confusion matrix, without normalization

[[ 122 151 21]

[ 102 3639 80]

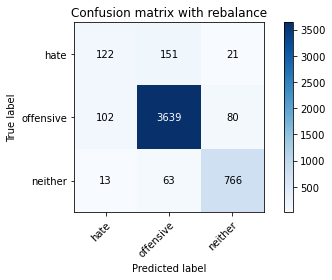
[ 13 63 766]]

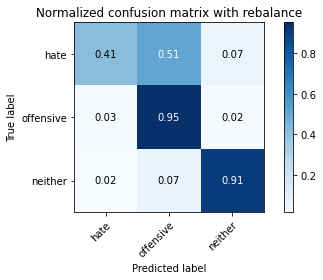
Normalized confusion matrix

[[0.41496599 0.51360544 0.07142857]

[0.02669458 0.95236849 0.02093693]

[0.01543943 0.07482185 0.90973872]]





precision recall f1-score support

hate 0.51 0.41 0.46 294

offensive 0.94 0.95 0.95 3821

neither 0.88 0.91 0.90 842

accuracy 0.91 4957

macro avg 0.78 0.76 0.77 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.26122443161476705, accuracy: 0.9132539842646762

training 4 epoch...

843.1814384460449 seconds used......

4 training finished! average train loss: 0.09174799886768009

total clf loss: 92.16206812096061 total adv loss: 22.96106159539977

evaluating...

Confusion matrix, without normalization

[[ 154 125 15]

[ 196 3542 83]

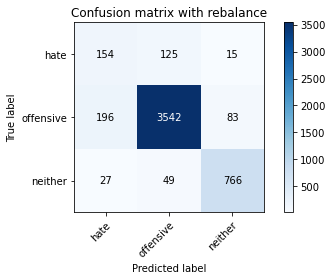
[ 27 49 766]]

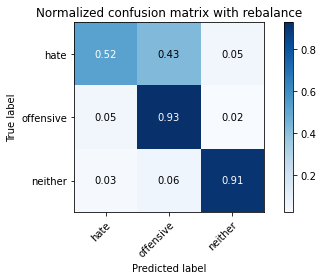
Normalized confusion matrix

[[0.52380952 0.42517007 0.05102041]

[0.05129547 0.92698247 0.02172206]

[0.03206651 0.05819477 0.90973872]]





precision recall f1-score support

hate 0.41 0.52 0.46 294

offensive 0.95 0.93 0.94 3821

neither 0.89 0.91 0.90 842

accuracy 0.90 4957

macro avg 0.75 0.79 0.77 4957

weighted avg 0.91 0.90 0.90 4957

average eval\_loss: 0.3556046359228804, accuracy: 0.9001412144442204

training 5 epoch...

843.5421345233917 seconds used......

5 training finished! average train loss: 0.04957690612062863

total clf loss: 49.85453808346938 total adv loss: 13.48494260055827

evaluating...

Confusion matrix, without normalization

[[ 89 191 14]

[ 86 3696 39]

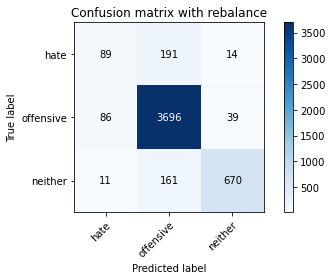
[ 11 161 670]]

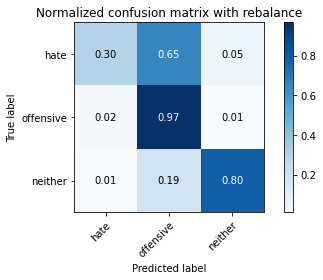
Normalized confusion matrix

[[0.30272109 0.64965986 0.04761905]

[0.0225072 0.96728605 0.01020675]

[0.01306413 0.1912114 0.79572447]]





precision recall f1-score support

hate 0.48 0.30 0.37 294

offensive 0.91 0.97 0.94 3821

neither 0.93 0.80 0.86 842

accuracy 0.90 4957

macro avg 0.77 0.69 0.72 4957

weighted avg 0.89 0.90 0.89 4957

average eval\_loss: 0.44706961297044706, accuracy: 0.8987290700020173

training 6 epoch...

843.5495235919952 seconds used......

6 training finished! average train loss: 0.03229455131249339

total clf loss: 32.56571175386489 total adv loss: 10.590335044279513

evaluating...

Confusion matrix, without normalization

[[ 123 152 19]

[ 106 3637 78]

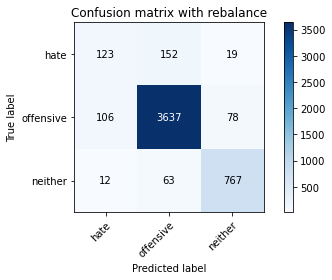
[ 12 63 767]]

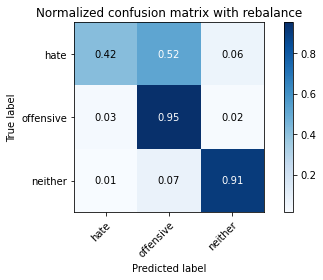
Normalized confusion matrix

[[0.41836735 0.5170068 0.06462585]

[0.02774143 0.95184507 0.0204135 ]

[0.01425178 0.07482185 0.91092637]]





precision recall f1-score support

hate 0.51 0.42 0.46 294

offensive 0.94 0.95 0.95 3821

neither 0.89 0.91 0.90 842

accuracy 0.91 4957

macro avg 0.78 0.76 0.77 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.47433364496286995, accuracy: 0.9132539842646762

training 7 epoch...

843.689049243927 seconds used......

7 training finished! average train loss: 0.01962402945804159

total clf loss: 19.841967420332367 total adv loss: 7.49860419684758

evaluating...

Confusion matrix, without normalization

[[ 100 176 18]

[ 94 3663 64]

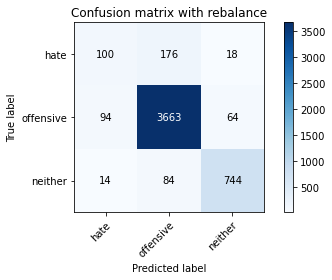
[ 14 84 744]]

Normalized confusion matrix

[[0.34013605 0.59863946 0.06122449]

[0.02460089 0.95864957 0.01674954]

[0.01662708 0.09976247 0.88361045]]





precision recall f1-score support

hate 0.48 0.34 0.40 294

offensive 0.93 0.96 0.95 3821

neither 0.90 0.88 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.73 0.75 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.4862969147747971, accuracy: 0.9092192858583821

training 8 epoch...

844.1287634372711 seconds used......

8 training finished! average train loss: 0.013212759478194365

total clf loss: 13.290544121580751 total adv loss: 3.6697346760182885

evaluating...

Confusion matrix, without normalization

[[ 107 165 22]

[ 136 3592 93]

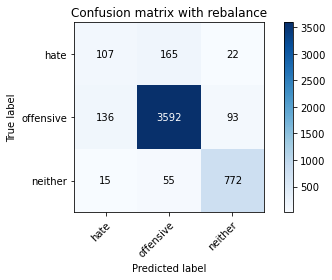
[ 15 55 772]]

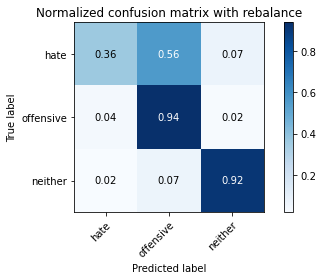
Normalized confusion matrix

[[0.36394558 0.56122449 0.07482993]

[0.03559278 0.94006805 0.02433918]

[0.01781473 0.06532067 0.91686461]]





precision recall f1-score support

hate 0.41 0.36 0.39 294

offensive 0.94 0.94 0.94 3821

neither 0.87 0.92 0.89 842

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: 0.5713009091191745, accuracy: 0.9019568287270526

training 9 epoch...

843.8624737262726 seconds used......

9 training finished! average train loss: 0.01698410373020095

total clf loss: 17.23963036136047 total adv loss: 7.827988169523048

evaluating...

Confusion matrix, without normalization

[[ 97 181 16]

[ 99 3666 56]

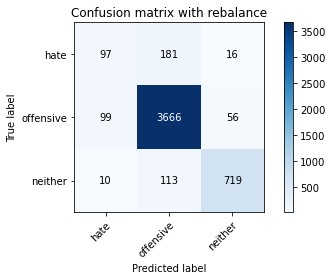
[ 10 113 719]]

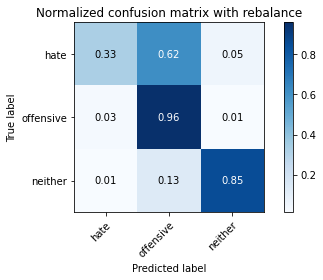
Normalized confusion matrix

[[0.32993197 0.61564626 0.05442177]

[0.02590945 0.9594347 0.01465585]

[0.01187648 0.13420428 0.85391924]]





precision recall f1-score support

hate 0.47 0.33 0.39 294

offensive 0.93 0.96 0.94 3821

neither 0.91 0.85 0.88 842

accuracy 0.90 4957

macro avg 0.77 0.71 0.74 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.5447379556738181, accuracy: 0.9041759128505145

training 10 epoch...

843.6688179969788 seconds used......

10 training finished! average train loss: 0.012323293956008514

total clf loss: 12.631645689918514 total adv loss: 8.138762381551409

evaluating...

Confusion matrix, without normalization

[[ 117 158 19]

[ 121 3620 80]

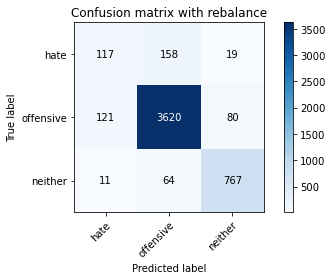
[ 11 64 767]]

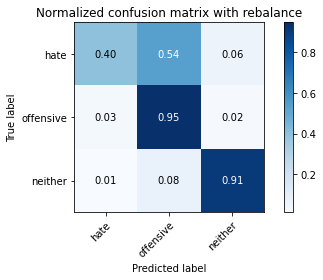
Normalized confusion matrix

[[0.39795918 0.53741497 0.06462585]

[0.0316671 0.94739597 0.02093693]

[0.01306413 0.0760095 0.91092637]]





precision recall f1-score support

hate 0.47 0.40 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 0.77 0.75 0.76 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.5523379209035012, accuracy: 0.908614081097438

2 run......

training 1 epoch...

843.4808235168457 seconds used......

1 training finished! average train loss: 0.2784871923053757

total clf loss: 283.11262047849596 total adv loss: 137.06650880811503

evaluating...

Confusion matrix, without normalization

[[ 53 211 30]

[ 49 3692 80]

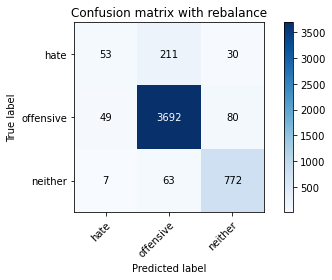
[ 7 63 772]]

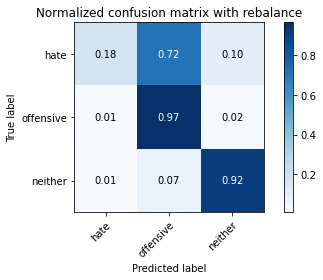
Normalized confusion matrix

[[0.18027211 0.71768707 0.10204082]

[0.01282387 0.9662392 0.02093693]

[0.00831354 0.07482185 0.91686461]]





precision recall f1-score support

hate 0.49 0.18 0.26 294

offensive 0.93 0.97 0.95 3821

neither 0.88 0.92 0.90 842

accuracy 0.91 4957

macro avg 0.76 0.69 0.70 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.2388689929183813, accuracy: 0.9112366350615292

training 2 epoch...

843.3779909610748 seconds used......

2 training finished! average train loss: 0.2009203161400247

total clf loss: 203.4471731716767 total adv loss: 82.68439144748845

evaluating...

Confusion matrix, without normalization

[[ 129 147 18]

[ 120 3615 86]

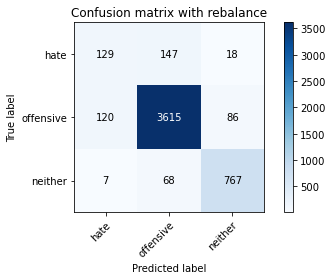
[ 7 68 767]]

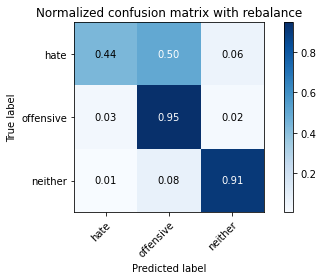
Normalized confusion matrix

[[0.43877551 0.5 0.06122449]

[0.03140539 0.94608741 0.0225072 ]

[0.00831354 0.0807601 0.91092637]]





precision recall f1-score support

hate 0.50 0.44 0.47 294

offensive 0.94 0.95 0.94 3821

neither 0.88 0.91 0.90 842

accuracy 0.91 4957

macro avg 0.78 0.77 0.77 4957

weighted avg 0.91 0.91 0.91 4957

average eval\_loss: 0.24496353880291985, accuracy: 0.910026225539641

training 3 epoch...

844.7255001068115 seconds used......

3 training finished! average train loss: 0.142132479643483

total clf loss: 143.1299871141091 total adv loss: 42.69134829358518

evaluating...

Confusion matrix, without normalization

[[ 75 192 27]

[ 59 3668 94]

[ 2 61 779]]

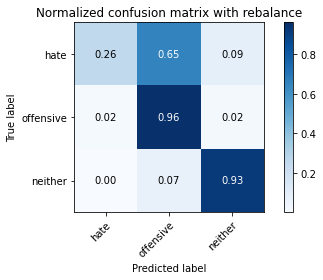
Normalized confusion matrix

[[0.25510204 0.65306122 0.09183673]

[0.01544098 0.95995813 0.02460089]

[0.0023753 0.07244656 0.92517815]]





precision recall f1-score support

hate 0.55 0.26 0.35 294

offensive 0.94 0.96 0.95 3821

neither 0.87 0.93 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.71 0.73 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.29085115406515766, accuracy: 0.9122453096631027

training 4 epoch...

843.8688905239105 seconds used......

4 training finished! average train loss: 0.09350859579236355

total clf loss: 93.93777670385316 total adv loss: 23.544990534737735

evaluating...

Confusion matrix, without normalization

[[ 112 167 15]

[ 100 3659 62]

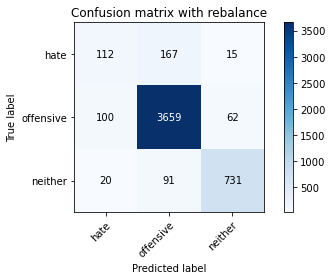
[ 20 91 731]]

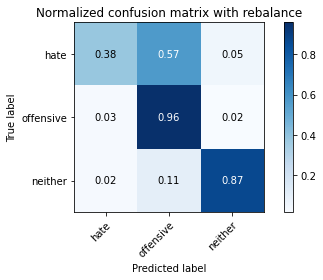
Normalized confusion matrix

[[0.38095238 0.56802721 0.05102041]

[0.02617116 0.95760272 0.01622612]

[0.02375297 0.10807601 0.86817102]]





precision recall f1-score support

hate 0.48 0.38 0.43 294

offensive 0.93 0.96 0.95 3821

neither 0.90 0.87 0.89 842

accuracy 0.91 4957

macro avg 0.77 0.74 0.75 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.3058612294689444, accuracy: 0.9082106112568086

training 5 epoch...

843.758882522583 seconds used......

5 training finished! average train loss: 0.05014235898908091

total clf loss: 50.58108391330461 total adv loss: 16.797277551889835

evaluating...

Confusion matrix, without normalization

[[ 118 158 18]

[ 113 3629 79]

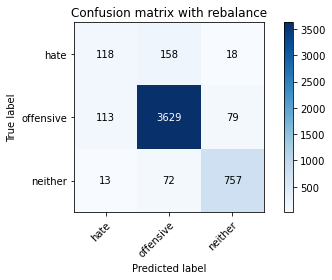
[ 13 72 757]]

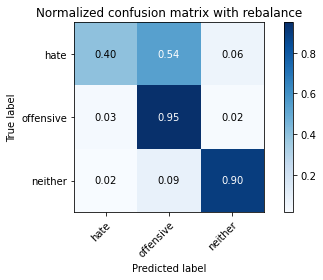
Normalized confusion matrix

[[0.40136054 0.53741497 0.06122449]

[0.02957341 0.94975137 0.02067522]

[0.01543943 0.08551069 0.89904988]]





precision recall f1-score support

hate 0.48 0.40 0.44 294

offensive 0.94 0.95 0.95 3821

neither 0.89 0.90 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: 0.3761625998224425, accuracy: 0.908614081097438

training 6 epoch...

1 run......

Downloading: 100%

411/411 [00:00<00:00, 13.1kB/s]

Downloading: 100%

263M/263M [00:09<00:00, 30.7MB/s]

training 1 epoch...

839.2627129554749 seconds used......

1 training finished! average train loss: 0.28081042487399593

total clf loss: 285.5930077303201 total adv loss: 140.58132397616282

evaluating...

Confusion matrix, without normalization

[[ 72 174 48]

[ 58 3627 136]

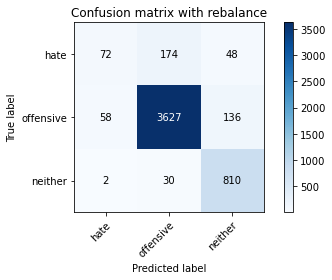
[ 2 30 810]]

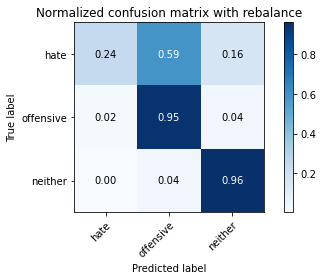
Normalized confusion matrix

[[0.24489796 0.59183673 0.16326531]

[0.01517927 0.94922795 0.03559278]

[0.0023753 0.03562945 0.96199525]]





precision recall f1-score support

hate 0.55 0.24 0.34 294

offensive 0.95 0.95 0.95 3821

neither 0.81 0.96 0.88 842

accuracy 0.91 4957

macro avg 0.77 0.72 0.72 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.26305253014734553, accuracy: 0.9096227556990115

training 2 epoch...

838.62180519104 seconds used......

2 training finished! average train loss: 0.2015947726805837

total clf loss: 204.19366446044296 total adv loss: 84.2329989617283

evaluating...

Confusion matrix, without normalization

[[ 69 193 32]

[ 46 3681 94]

[ 0 56 786]]

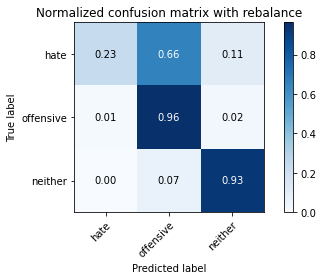
Normalized confusion matrix

[[0.23469388 0.65646259 0.10884354]

[0.01203873 0.96336038 0.02460089]

[0. 0.06650831 0.93349169]]





precision recall f1-score support

hate 0.60 0.23 0.34 294

offensive 0.94 0.96 0.95 3821

neither 0.86 0.93 0.90 842

accuracy 0.92 4957

macro avg 0.80 0.71 0.73 4957

weighted avg 0.90 0.92 0.90 4957

average eval\_loss: 0.2394227592514888, accuracy: 0.9150695985475086

training 3 epoch...

838.963947057724 seconds used......

3 training finished! average train loss: 0.13975217208808904

total clf loss: 140.99494017846882 total adv loss: 47.215711534885486

evaluating...

Confusion matrix, without normalization

[[ 94 180 20]

[ 87 3638 96]

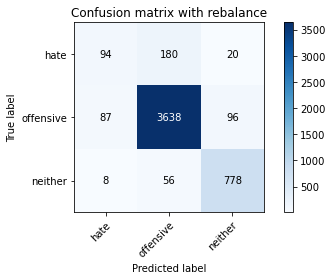
[ 8 56 778]]

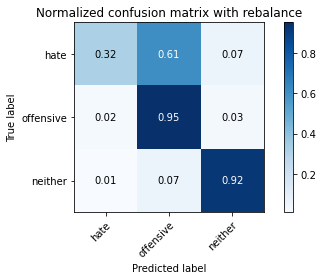
Normalized confusion matrix

[[0.31972789 0.6122449 0.06802721]

[0.02276891 0.95210678 0.02512431]

[0.00950119 0.06650831 0.9239905 ]]





precision recall f1-score support

hate 0.50 0.32 0.39 294

offensive 0.94 0.95 0.95 3821

neither 0.87 0.92 0.90 842

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: 0.2726056397396831, accuracy: 0.9098244906193262

training 4 epoch...

839.0645432472229 seconds used......

4 training finished! average train loss: 0.0857400554130788

total clf loss: 86.3286870099837 total adv loss: 25.491042164740065

evaluating...

Confusion matrix, without normalization

[[ 103 172 19]

[ 116 3640 65]

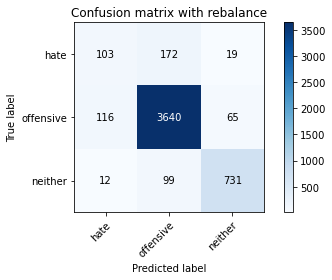
[ 12 99 731]]

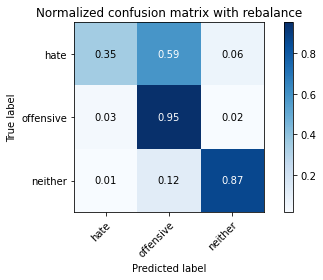
Normalized confusion matrix

[[0.35034014 0.58503401 0.06462585]

[0.03035854 0.9526302 0.01701125]

[0.01425178 0.1175772 0.86817102]]





precision recall f1-score support

hate 0.45 0.35 0.39 294

offensive 0.93 0.95 0.94 3821

neither 0.90 0.87 0.88 842

accuracy 0.90 4957

macro avg 0.76 0.72 0.74 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.33955312897582657, accuracy: 0.9025620334879968

training 5 epoch...

838.934531211853 seconds used......

5 training finished! average train loss: 0.04445314766889492

total clf loss: 44.836035439628176 total adv loss: 14.77026044186971

evaluating...

Confusion matrix, without normalization

[[ 98 182 14]

[ 89 3664 68]

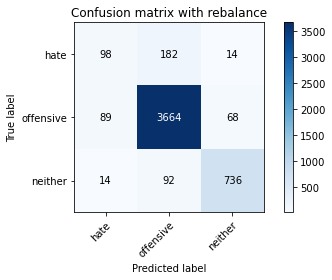
[ 14 92 736]]

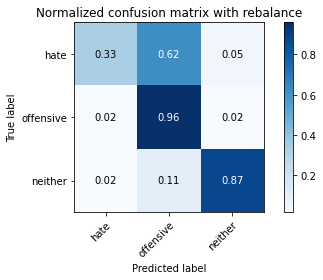
Normalized confusion matrix

[[0.33333333 0.61904762 0.04761905]

[0.02329233 0.95891128 0.01779639]

[0.01662708 0.10926366 0.87410926]]





precision recall f1-score support

hate 0.49 0.33 0.40 294

offensive 0.93 0.96 0.94 3821

neither 0.90 0.87 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: 0.4187093220149813, accuracy: 0.9074036715755497

training 6 epoch...

839.4248788356781 seconds used......

6 training finished! average train loss: 0.029971518069179446

total clf loss: 30.218931803960004 total adv loss: 9.743716854201125

evaluating...

Confusion matrix, without normalization

[[ 83 194 17]

[ 71 3675 75]

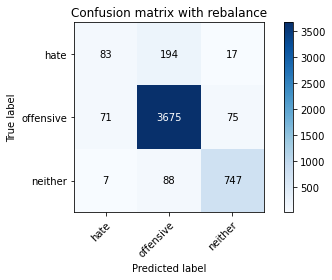
[ 7 88 747]]

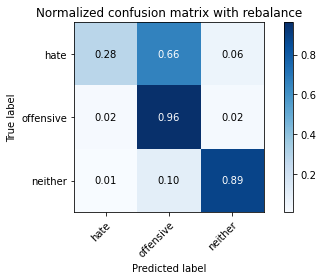
Normalized confusion matrix

[[0.28231293 0.65986395 0.05782313]

[0.01858152 0.96179011 0.01962837]

[0.00831354 0.10451306 0.8871734 ]]





precision recall f1-score support

hate 0.52 0.28 0.36 294

offensive 0.93 0.96 0.94 3821

neither 0.89 0.89 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.71 0.73 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.4911290629722991, accuracy: 0.9088158160177526

training 7 epoch...

839.2157702445984 seconds used......

7 training finished! average train loss: 0.01698209404256182

total clf loss: 17.279545558769314 total adv loss: 8.666165905584133

evaluating...

Confusion matrix, without normalization

[[ 99 175 20]

[ 117 3614 90]

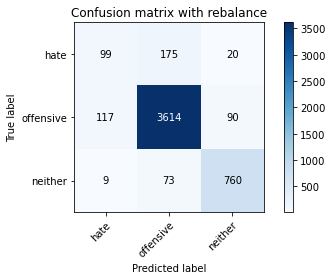
[ 9 73 760]]

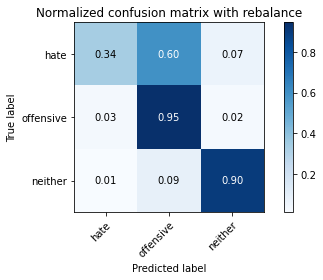
Normalized confusion matrix

[[0.33673469 0.5952381 0.06802721]

[0.03062026 0.9458257 0.02355404]

[0.01068884 0.08669834 0.90261283]]





precision recall f1-score support

hate 0.44 0.34 0.38 294

offensive 0.94 0.95 0.94 3821

neither 0.87 0.90 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: 0.5448203727843769, accuracy: 0.9023602985676821

training 8 epoch...

839.6730508804321 seconds used......

8 training finished! average train loss: 0.02103660470167538

total clf loss: 21.44001406364623 total adv loss: 11.434043663501136

evaluating...

Confusion matrix, without normalization

[[ 85 196 13]

[ 79 3690 52]

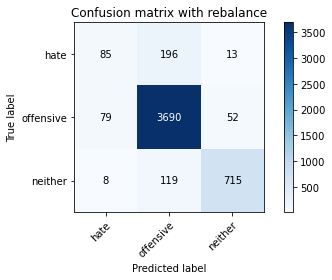
[ 8 119 715]]

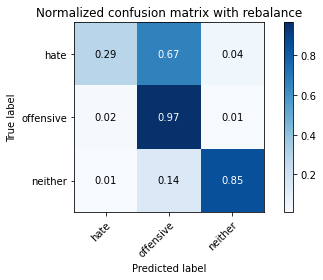
Normalized confusion matrix

[[0.28911565 0.66666667 0.04421769]

[0.02067522 0.96571578 0.013609 ]

[0.00950119 0.14133017 0.84916865]]





precision recall f1-score support

hate 0.49 0.29 0.36 294

offensive 0.92 0.97 0.94 3821

neither 0.92 0.85 0.88 842

accuracy 0.91 4957

macro avg 0.78 0.70 0.73 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.541119558686073, accuracy: 0.9057897922130321

training 9 epoch...

839.1734800338745 seconds used......

9 training finished! average train loss: 0.013941016837630712

total clf loss: 14.216868623272603 total adv loss: 7.747598390910689

evaluating...

Confusion matrix, without normalization

[[ 99 181 14]

[ 115 3637 69]

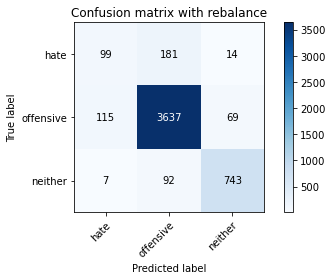
[ 7 92 743]]

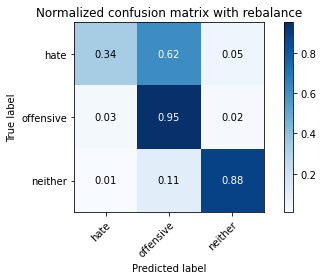
Normalized confusion matrix

[[0.33673469 0.61564626 0.04761905]

[0.03009683 0.95184507 0.0180581 ]

[0.00831354 0.10926366 0.8824228 ]]





precision recall f1-score support

hate 0.45 0.34 0.38 294

offensive 0.93 0.95 0.94 3821

neither 0.90 0.88 0.89 842

accuracy 0.90 4957

macro avg 0.76 0.72 0.74 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.5475615382700844, accuracy: 0.9035707080895703

training 10 epoch...

838.8585152626038 seconds used......

10 training finished! average train loss: 0.011112269735529099

total clf loss: 11.35465454434052 total adv loss: 6.625659416944024

evaluating...

Confusion matrix, without normalization

[[ 105 176 13]

[ 111 3652 58]

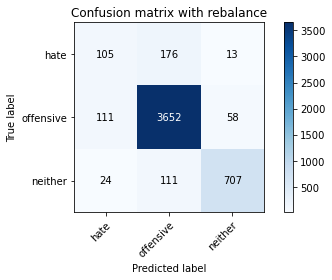
[ 24 111 707]]

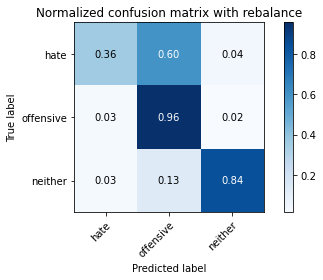
Normalized confusion matrix

[[0.35714286 0.59863946 0.04421769]

[0.02904999 0.95577074 0.01517927]

[0.02850356 0.13182898 0.83966746]]





precision recall f1-score support

hate 0.44 0.36 0.39 294

offensive 0.93 0.96 0.94 3821

neither 0.91 0.84 0.87 842

accuracy 0.90 4957

macro avg 0.76 0.72 0.74 4957

weighted avg 0.89 0.90 0.90 4957

average eval\_loss: 0.48591473859508577, accuracy: 0.9005446842848497

2 run......

training 1 epoch...

838.9333980083466 seconds used......

1 training finished! average train loss: 0.2764340340753176

total clf loss: 280.94613747578114 total adv loss: 134.47151508089155

evaluating...

Confusion matrix, without normalization

[[ 126 135 33]

[ 126 3595 100]

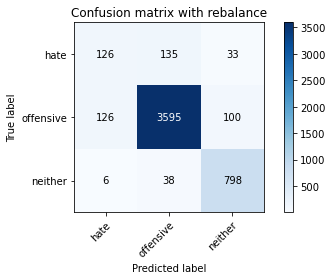
[ 6 38 798]]

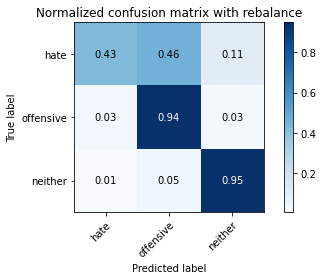
Normalized confusion matrix

[[0.42857143 0.45918367 0.1122449 ]

[0.03297566 0.94085318 0.02617116]

[0.00712589 0.04513064 0.94774347]]





precision recall f1-score support

hate 0.49 0.43 0.46 294

offensive 0.95 0.94 0.95 3821

neither 0.86 0.95 0.90 842

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: 0.24797191413780373, accuracy: 0.9116401049021585

training 2 epoch...

838.5830321311951 seconds used......

2 training finished! average train loss: 0.19834805517166387

total clf loss: 200.73725160304457 total adv loss: 79.51961480805767

evaluating...

Confusion matrix, without normalization

[[ 72 188 34]

[ 48 3656 117]

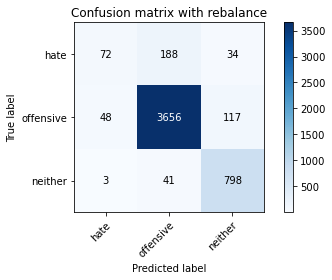
[ 3 41 798]]

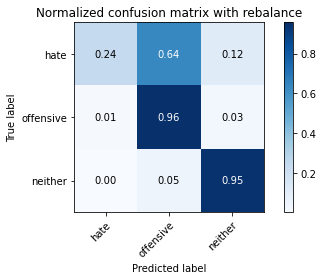
Normalized confusion matrix

[[0.24489796 0.63945578 0.11564626]

[0.01256216 0.95681759 0.03062026]

[0.00356295 0.04869359 0.94774347]]





precision recall f1-score support

hate 0.59 0.24 0.35 294

offensive 0.94 0.96 0.95 3821

neither 0.84 0.95 0.89 842

accuracy 0.91 4957

macro avg 0.79 0.72 0.73 4957

weighted avg 0.90 0.91 0.90 4957

average eval\_loss: 0.25731439079578605, accuracy: 0.9130522493443615

training 3 epoch...

838.5440537929535 seconds used......

3 training finished! average train loss: 0.1410556882173912

total clf loss: 142.13091850932688 total adv loss: 44.07351921764348

evaluating...

Confusion matrix, without normalization

[[ 105 169 20]

[ 88 3660 73]

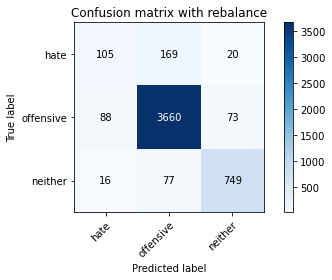
[ 16 77 749]]

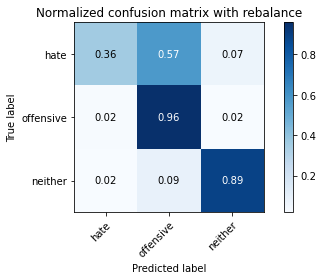
Normalized confusion matrix

[[0.35714286 0.57482993 0.06802721]

[0.02303062 0.95786443 0.01910495]

[0.01900238 0.09144893 0.88954869]]





precision recall f1-score support

hate 0.50 0.36 0.42 294

offensive 0.94 0.96 0.95 3821

neither 0.89 0.89 0.89 842

accuracy 0.91 4957

macro avg 0.78 0.73 0.75 4957

weighted avg 0.90 0.91 0.91 4957

average eval\_loss: 0.29152392286707784, accuracy: 0.910631430300585

training 4 epoch...

838.3518617153168 seconds used......

4 training finished! average train loss: 0.08763534350195949

total clf loss: 87.99132501686108 total adv loss: 21.14128642199512

evaluating...

Confusion matrix, without normalization

[[ 105 174 15]

[ 93 3669 59]

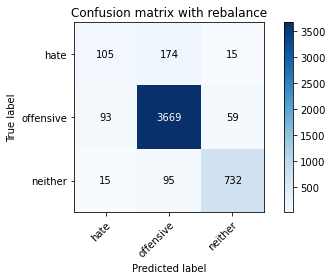
[ 15 95 732]]

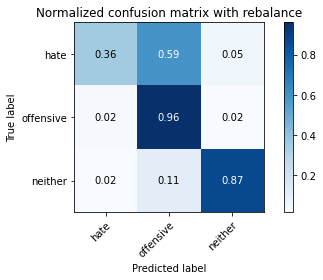
Normalized confusion matrix

[[0.35714286 0.59183673 0.05102041]

[0.02433918 0.96021984 0.01544098]

[0.01781473 0.1128266 0.86935867]]





precision recall f1-score support

hate 0.49 0.36 0.41 294

offensive 0.93 0.96 0.95 3821

neither 0.91 0.87 0.89 842

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: 0.34540582841756046, accuracy: 0.9090175509380674

training 5 epoch...

838.9938762187958 seconds used......

5 training finished! average train loss: 0.04819312459702844

total clf loss: 48.52600566903129 total adv loss: 14.36852208630377

evaluating...

Confusion matrix, without normalization

[[ 108 168 18]

[ 117 3634 70]

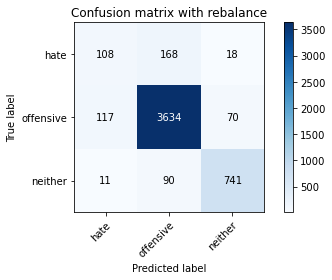
[ 11 90 741]]

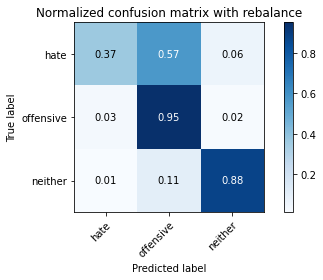
Normalized confusion matrix

[[0.36734694 0.57142857 0.06122449]

[0.03062026 0.95105993 0.01831981]

[0.01306413 0.10688836 0.88004751]]





precision recall f1-score support

hate 0.46 0.37 0.41 294

offensive 0.93 0.95 0.94 3821

neither 0.89 0.88 0.89 842

accuracy 0.90 4957

macro avg 0.76 0.73 0.75 4957

weighted avg 0.90 0.90 0.90 4957

average eval\_loss: 0.37768662680501885, accuracy: 0.9043776477708292

training 6 epoch...