alpha = 2

hate as no-offensive

training 1 epoch...

44.67928719520569 seconds used......

1 training finished! average train loss: 1.5162141155778317

total clf loss: 116.55722540616989 total adv loss: 71.3576937392354

evaluating...

Confusion matrix, without normalization

[[ 213 60 21]

[ 723 3032 66]

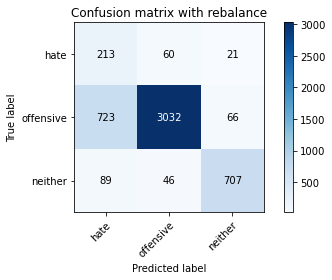
[ 89 46 707]]

Normalized confusion matrix

[[0.7244898 0.20408163 0.07142857]

[0.18921748 0.79350955 0.01727297]

[0.10570071 0.05463183 0.83966746]]





precision recall f1-score support

hate 0.21 0.72 0.32 294

offensive 0.97 0.79 0.87 3821

neither 0.89 0.84 0.86 842

accuracy 0.80 4957

macro avg 0.69 0.79 0.69 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.44217297440815356, accuracy: 0.79725640508372

training 2 epoch...

44.71063947677612 seconds used......

2 training finished! average train loss: 1.003700192734512

total clf loss: 71.7840847671032 total adv loss: 49.92432380467653

evaluating...

Confusion matrix, without normalization

[[ 204 61 29]

[ 638 3067 116]

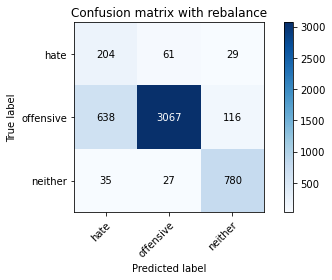
[ 35 27 780]]

Normalized confusion matrix

[[0.69387755 0.20748299 0.09863946]

[0.166972 0.80266946 0.03035854]

[0.0415677 0.03206651 0.9263658 ]]





precision recall f1-score support

hate 0.23 0.69 0.35 294

offensive 0.97 0.80 0.88 3821

neither 0.84 0.93 0.88 842

accuracy 0.82 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.42803618347933214, accuracy: 0.8172281621948759

training 3 epoch...

44.68569612503052 seconds used......

3 training finished! average train loss: 0.7495142200536895

total clf loss: 53.55930585414171 total adv loss: 37.30381275713444

evaluating...

Confusion matrix, without normalization

[[ 249 34 11]

[1183 2563 75]

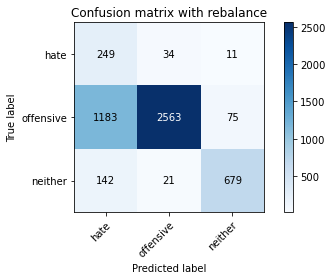
[ 142 21 679]]

Normalized confusion matrix

[[0.84693878 0.11564626 0.03741497]

[0.30960482 0.67076681 0.01962837]

[0.16864608 0.02494062 0.8064133 ]]





precision recall f1-score support

hate 0.16 0.85 0.27 294

offensive 0.98 0.67 0.80 3821

neither 0.89 0.81 0.85 842

accuracy 0.70 4957

macro avg 0.67 0.77 0.64 4957

weighted avg 0.91 0.70 0.77 4957

average eval\_loss: 0.6758690741393836, accuracy: 0.7042566068186403

training 4 epoch...

44.68309473991394 seconds used......

4 training finished! average train loss: 0.49554105323047665

total clf loss: 34.218011155724525 total adv loss: 25.25975442165509

evaluating...

Confusion matrix, without normalization

[[ 210 64 20]

[ 578 3127 116]

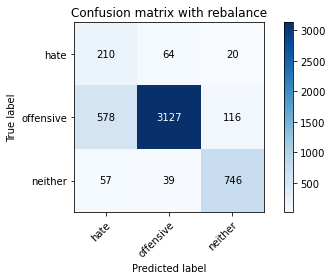
[ 57 39 746]]

Normalized confusion matrix

[[0.71428571 0.21768707 0.06802721]

[0.1512693 0.81837215 0.03035854]

[0.06769596 0.04631829 0.88598575]]





precision recall f1-score support

hate 0.25 0.71 0.37 294

offensive 0.97 0.82 0.89 3821

neither 0.85 0.89 0.87 842

accuracy 0.82 4957

macro avg 0.69 0.81 0.71 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.48975786719951897, accuracy: 0.8236836796449465

training 5 epoch...

44.67363429069519 seconds used......

5 training finished! average train loss: 0.27419801474663247

total clf loss: 19.33719136286527 total adv loss: 13.77533452049829

evaluating...

Confusion matrix, without normalization

[[ 216 61 17]

[ 621 3084 116]

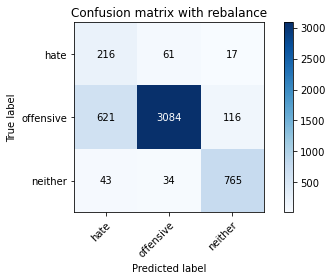
[ 43 34 765]]

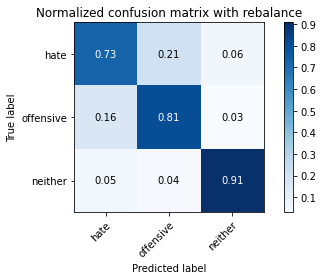
Normalized confusion matrix

[[0.73469388 0.20748299 0.05782313]

[0.1625229 0.80711856 0.03035854]

[0.05106888 0.04038005 0.90855107]]





precision recall f1-score support

hate 0.25 0.73 0.37 294

offensive 0.97 0.81 0.88 3821

neither 0.85 0.91 0.88 842

accuracy 0.82 4957

macro avg 0.69 0.82 0.71 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.5789205358709179, accuracy: 0.8200524510792818

training 6 epoch...

44.69810628890991 seconds used......

6 training finished! average train loss: 0.1843346968423902

total clf loss: 12.104279518360272 total adv loss: 9.708476846150006

evaluating...

Confusion matrix, without normalization

[[ 207 63 24]

[ 656 3051 114]

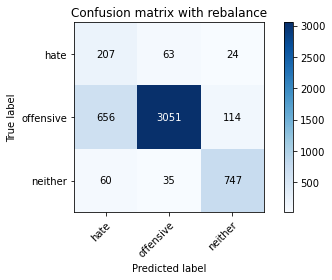
[ 60 35 747]]

Normalized confusion matrix

[[0.70408163 0.21428571 0.08163265]

[0.17168281 0.79848207 0.02983512]

[0.07125891 0.0415677 0.8871734 ]]





precision recall f1-score support

hate 0.22 0.70 0.34 294

offensive 0.97 0.80 0.88 3821

neither 0.84 0.89 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.6999070711674229, accuracy: 0.8079483558603995

training 7 epoch...

44.68305587768555 seconds used......

7 training finished! average train loss: 0.10185639572068396

total clf loss: 6.545645361300558 total adv loss: 5.435899155825609

evaluating...

Confusion matrix, without normalization

[[ 221 50 23]

[ 730 2939 152]

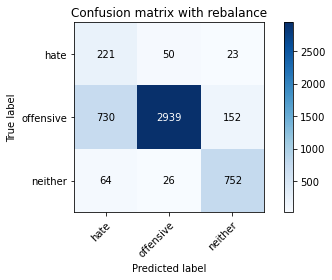
[ 64 26 752]]

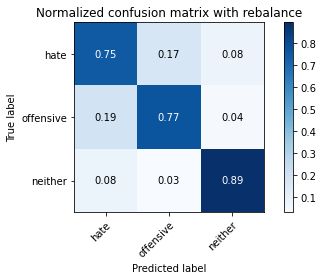
Normalized confusion matrix

[[0.75170068 0.17006803 0.07823129]

[0.19104946 0.76917037 0.03978016]

[0.0760095 0.03087886 0.89311164]]





precision recall f1-score support

hate 0.22 0.75 0.34 294

offensive 0.97 0.77 0.86 3821

neither 0.81 0.89 0.85 842

accuracy 0.79 4957

macro avg 0.67 0.80 0.68 4957

weighted avg 0.90 0.79 0.83 4957

average eval\_loss: 0.8852929377650481, accuracy: 0.7891870082711318

training 8 epoch...

44.7274112701416 seconds used......

8 training finished! average train loss: 0.0700784612546419

total clf loss: 5.187768459320068 total adv loss: 3.3978241751165115

evaluating...

Confusion matrix, without normalization

[[ 218 52 24]

[ 769 2900 152]

[ 52 27 763]]

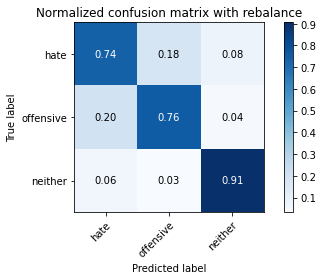
Normalized confusion matrix

[[0.7414966 0.17687075 0.08163265]

[0.20125622 0.75896362 0.03978016]

[0.06175772 0.03206651 0.90617577]]





precision recall f1-score support

hate 0.21 0.74 0.33 294

offensive 0.97 0.76 0.85 3821

neither 0.81 0.91 0.86 842

accuracy 0.78 4957

macro avg 0.67 0.80 0.68 4957

weighted avg 0.90 0.78 0.82 4957

average eval\_loss: 1.0284402715267553, accuracy: 0.7829332257413758

training 9 epoch...

44.65260100364685 seconds used......

9 training finished! average train loss: 0.08741037973219104

total clf loss: 5.7996501726447605 total adv loss: 4.573762395142694

evaluating...

Confusion matrix, without normalization

[[ 231 47 16]

[ 906 2839 76]

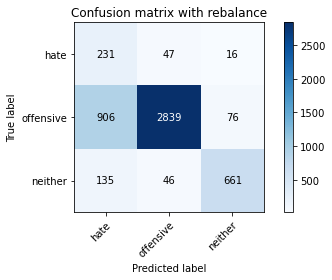
[ 135 46 661]]

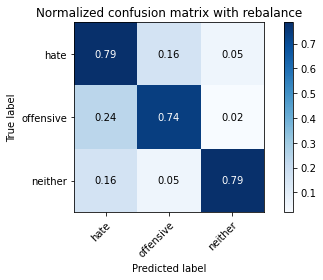
Normalized confusion matrix

[[0.78571429 0.15986395 0.05442177]

[0.2371107 0.74299921 0.01989008]

[0.16033254 0.05463183 0.78503563]]





precision recall f1-score support

hate 0.18 0.79 0.30 294

offensive 0.97 0.74 0.84 3821

neither 0.88 0.79 0.83 842

accuracy 0.75 4957

macro avg 0.68 0.77 0.65 4957

weighted avg 0.91 0.75 0.81 4957

average eval\_loss: 1.168935737737845, accuracy: 0.7526729876941699

training 10 epoch...

44.68110251426697 seconds used......

10 training finished! average train loss: 0.053291254097865406

total clf loss: 3.681333867192734 total adv loss: 2.715735313012374

evaluating...

Confusion matrix, without normalization

[[ 224 50 20]

[ 801 2860 160]

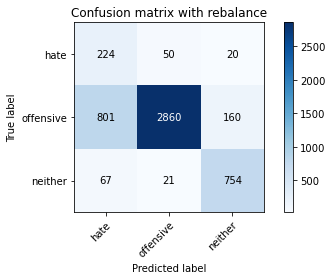
[ 67 21 754]]

Normalized confusion matrix

[[0.76190476 0.17006803 0.06802721]

[0.20963099 0.74849516 0.04187386]

[0.07957245 0.02494062 0.89548694]]





precision recall f1-score support

hate 0.21 0.76 0.32 294

offensive 0.98 0.75 0.85 3821

neither 0.81 0.90 0.85 842

accuracy 0.77 4957

macro avg 0.66 0.80 0.67 4957

weighted avg 0.90 0.77 0.82 4957

average eval\_loss: 1.1941020575336991, accuracy: 0.7742586241678434

2 run

training 1 epoch...

44.893216609954834 seconds used......

1 training finished! average train loss: 1.5383013270751775

total clf loss: 119.68751826882362 total adv loss: 71.68100465834141

evaluating...

Confusion matrix, without normalization

[[ 223 39 32]

[ 865 2812 144]

[ 58 12 772]]

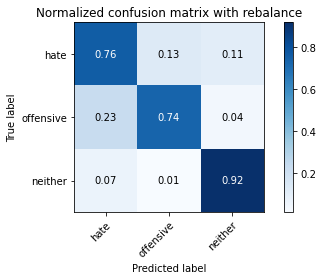
Normalized confusion matrix

[[0.7585034 0.13265306 0.10884354]

[0.22638053 0.735933 0.03768647]

[0.06888361 0.01425178 0.91686461]]





precision recall f1-score support

hate 0.19 0.76 0.31 294

offensive 0.98 0.74 0.84 3821

neither 0.81 0.92 0.86 842

accuracy 0.77 4957

macro avg 0.66 0.80 0.67 4957

weighted avg 0.91 0.77 0.81 4957

average eval\_loss: 0.569561761894053, accuracy: 0.7680048416380876

training 2 epoch...

44.82947826385498 seconds used......

2 training finished! average train loss: 1.0351820743223379

total clf loss: 72.72822196036577 total adv loss: 52.14395621418953

evaluating...

Confusion matrix, without normalization

[[ 243 39 12]

[1013 2721 87]

[ 107 20 715]]

Normalized confusion matrix

[[0.82653061 0.13265306 0.04081633]

[0.26511384 0.71211725 0.02276891]

[0.12707838 0.02375297 0.84916865]]





precision recall f1-score support

hate 0.18 0.83 0.29 294

offensive 0.98 0.71 0.82 3821

neither 0.88 0.85 0.86 842

accuracy 0.74 4957

macro avg 0.68 0.80 0.66 4957

weighted avg 0.91 0.74 0.80 4957

average eval\_loss: 0.6185475915910736, accuracy: 0.7421827718378051

training 3 epoch...

44.87274885177612 seconds used......

3 training finished! average train loss: 0.7737668567930746

total clf loss: 53.11498290300369 total adv loss: 39.59957496263087

evaluating...

Confusion matrix, without normalization

[[ 212 50 32]

[ 714 2961 146]

[ 30 22 790]]

Normalized confusion matrix

[[0.72108844 0.17006803 0.10884354]

[0.18686208 0.77492803 0.03820989]

[0.03562945 0.02612827 0.93824228]]





precision recall f1-score support

hate 0.22 0.72 0.34 294

offensive 0.98 0.77 0.86 3821

neither 0.82 0.94 0.87 842

accuracy 0.80 4957

macro avg 0.67 0.81 0.69 4957

weighted avg 0.90 0.80 0.83 4957

average eval\_loss: 0.49525150662708667, accuracy: 0.7994754892071818

training 4 epoch...

44.93176794052124 seconds used......

4 training finished! average train loss: 0.47506592967356853

total clf loss: 32.51260966062546 total adv loss: 24.361832212656736

evaluating...

Confusion matrix, without normalization

[[ 239 44 11]

[ 890 2862 69]

[ 107 48 687]]

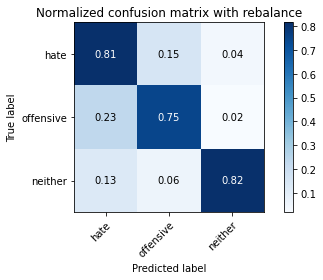
Normalized confusion matrix

[[0.81292517 0.14965986 0.03741497]

[0.23292332 0.74901858 0.0180581 ]

[0.12707838 0.05700713 0.81591449]]





precision recall f1-score support

hate 0.19 0.81 0.31 294

offensive 0.97 0.75 0.84 3821

neither 0.90 0.82 0.85 842

accuracy 0.76 4957

macro avg 0.69 0.79 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 0.6248272371688678, accuracy: 0.7641718781521081

training 5 epoch...

44.82429790496826 seconds used......

5 training finished! average train loss: 0.27599821197228475

total clf loss: 18.672781077213585 total adv loss: 14.261456669075415

evaluating...

Confusion matrix, without normalization

[[ 215 59 20]

[ 712 2988 121]

[ 60 36 746]]

Normalized confusion matrix

[[0.73129252 0.20068027 0.06802721]

[0.18633865 0.78199424 0.0316671 ]

[0.07125891 0.04275534 0.88598575]]





precision recall f1-score support

hate 0.22 0.73 0.34 294

offensive 0.97 0.78 0.87 3821

neither 0.84 0.89 0.86 842

accuracy 0.80 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.90 0.80 0.83 4957

average eval\_loss: 0.6643962036037157, accuracy: 0.7966512003227759

training 6 epoch...

44.810418128967285 seconds used......

6 training finished! average train loss: 0.14589244796763778

total clf loss: 10.443353488110006 total adv loss: 7.252127544241375

evaluating...

Confusion matrix, without normalization

[[ 218 52 24]

[ 699 2983 139]

[ 51 40 751]]

Normalized confusion matrix

[[0.7414966 0.17687075 0.08163265]

[0.1829364 0.78068568 0.03637791]

[0.06057007 0.04750594 0.89192399]]





precision recall f1-score support

hate 0.23 0.74 0.35 294

offensive 0.97 0.78 0.87 3821

neither 0.82 0.89 0.86 842

accuracy 0.80 4957

macro avg 0.67 0.80 0.69 4957

weighted avg 0.90 0.80 0.83 4957

average eval\_loss: 0.8132454527526973, accuracy: 0.79725640508372

training 7 epoch...

44.79766321182251 seconds used......

7 training finished! average train loss: 0.10240752244714038

total clf loss: 7.130032268818468 total adv loss: 5.190827057920615

evaluating...

Confusion matrix, without normalization

[[ 229 50 15]

[ 956 2750 115]

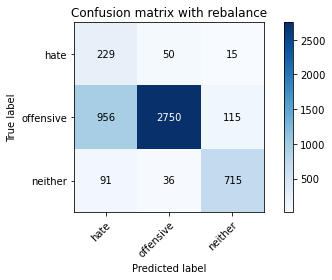
[ 91 36 715]]

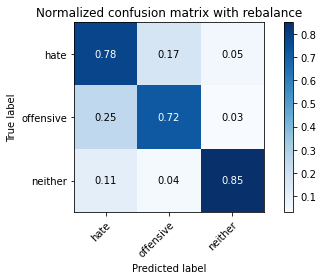
Normalized confusion matrix

[[0.77891156 0.17006803 0.05102041]

[0.25019628 0.71970688 0.03009683]

[0.10807601 0.04275534 0.84916865]]





precision recall f1-score support

hate 0.18 0.78 0.29 294

offensive 0.97 0.72 0.83 3821

neither 0.85 0.85 0.85 842

accuracy 0.75 4957

macro avg 0.67 0.78 0.66 4957

weighted avg 0.90 0.75 0.80 4957

average eval\_loss: 1.05378219869829, accuracy: 0.7452087956425257

training 8 epoch...

44.857351541519165 seconds used......

8 training finished! average train loss: 0.06006372884005766

total clf loss: 4.291158331790939 total adv loss: 2.989869652006746

evaluating...

Confusion matrix, without normalization

[[ 228 52 14]

[ 782 2921 118]

[ 97 38 707]]

Normalized confusion matrix

[[0.7755102 0.17687075 0.04761905]

[0.20465847 0.76445957 0.03088197]

[0.1152019 0.04513064 0.83966746]]





precision recall f1-score support

hate 0.21 0.78 0.33 294

offensive 0.97 0.76 0.86 3821

neither 0.84 0.84 0.84 842

accuracy 0.78 4957

macro avg 0.67 0.79 0.67 4957

weighted avg 0.90 0.78 0.82 4957

average eval\_loss: 1.0544286428485066, accuracy: 0.7778898527335082

training 9 epoch...

44.80414962768555 seconds used......

9 training finished! average train loss: 0.06281246670538135

total clf loss: 4.391867569065653 total adv loss: 3.174532148925209

evaluating...

Confusion matrix, without normalization

[[ 217 63 14]

[ 568 3177 76]

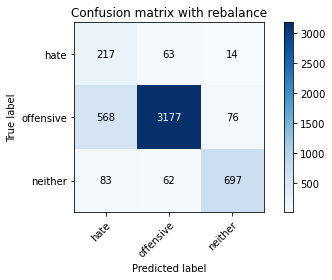
[ 83 62 697]]

Normalized confusion matrix

[[0.73809524 0.21428571 0.04761905]

[0.14865219 0.83145773 0.01989008]

[0.09857482 0.0736342 0.82779097]]





precision recall f1-score support

hate 0.25 0.74 0.37 294

offensive 0.96 0.83 0.89 3821

neither 0.89 0.83 0.86 842

accuracy 0.83 4957

macro avg 0.70 0.80 0.71 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.8138925323146395, accuracy: 0.8252975590074642

training 10 epoch...

44.801703691482544 seconds used......

10 training finished! average train loss: 0.06994320809611729

total clf loss: 4.1244515962316655 total adv loss: 3.9179184623717447

evaluating...

Confusion matrix, without normalization

[[ 225 48 21]

[ 850 2837 134]

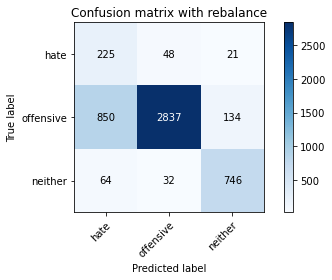
[ 64 32 746]]

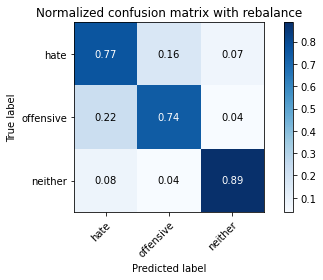
Normalized confusion matrix

[[0.76530612 0.16326531 0.07142857]

[0.22245485 0.74247579 0.03506935]

[0.0760095 0.03800475 0.88598575]]





precision recall f1-score support

hate 0.20 0.77 0.31 294

offensive 0.97 0.74 0.84 3821

neither 0.83 0.89 0.86 842

accuracy 0.77 4957

macro avg 0.67 0.80 0.67 4957

weighted avg 0.90 0.77 0.81 4957

average eval\_loss: 1.142670659183134, accuracy: 0.7682065765584023

3 run

training 1 epoch...

44.754552602767944 seconds used......

1 training finished! average train loss: 1.5721581217489744

total clf loss: 121.83338090777397 total adv loss: 73.50282913446426

evaluating...

Confusion matrix, without normalization

[[ 188 57 49]

[ 650 3015 156]

[ 28 20 794]]

Normalized confusion matrix

[[0.63945578 0.19387755 0.16666667]

[0.17011254 0.78906046 0.04082701]

[0.03325416 0.02375297 0.94299287]]





precision recall f1-score support

hate 0.22 0.64 0.32 294

offensive 0.98 0.79 0.87 3821

neither 0.79 0.94 0.86 842

accuracy 0.81 4957

macro avg 0.66 0.79 0.69 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.4666076128280932, accuracy: 0.8063344764978818

training 2 epoch...

44.669628381729126 seconds used......

2 training finished! average train loss: 1.0341458651754591

total clf loss: 74.44499458372593 total adv loss: 51.19697384163737

evaluating...

Confusion matrix, without normalization

[[ 229 43 22]

[ 743 2976 102]

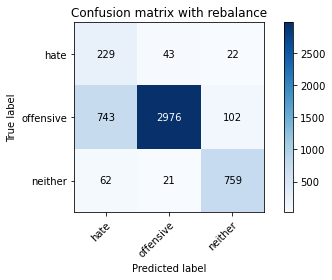
[ 62 21 759]]

Normalized confusion matrix

[[0.77891156 0.1462585 0.07482993]

[0.19445171 0.7788537 0.02669458]

[0.0736342 0.02494062 0.90142518]]





precision recall f1-score support

hate 0.22 0.78 0.34 294

offensive 0.98 0.78 0.87 3821

neither 0.86 0.90 0.88 842

accuracy 0.80 4957

macro avg 0.69 0.82 0.70 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.47460496458675594, accuracy: 0.7996772241274964

training 3 epoch...

44.75868082046509 seconds used......

3 training finished! average train loss: 0.7422668928204225

total clf loss: 51.57908146455884 total adv loss: 37.67427842132747

evaluating...

Confusion matrix, without normalization

[[ 233 44 17]

[ 843 2864 114]

[ 81 23 738]]

Normalized confusion matrix

[[0.79251701 0.14965986 0.05782313]

[0.22062287 0.749542 0.02983512]

[0.09619952 0.02731591 0.87648456]]





precision recall f1-score support

hate 0.20 0.79 0.32 294

offensive 0.98 0.75 0.85 3821

neither 0.85 0.88 0.86 842

accuracy 0.77 4957

macro avg 0.68 0.81 0.68 4957

weighted avg 0.91 0.77 0.82 4957

average eval\_loss: 0.5540658447500919, accuracy: 0.7736534194068994

training 4 epoch...

44.675995111465454 seconds used......

4 training finished! average train loss: 0.4408051505554141

total clf loss: 30.942604763433337 total adv loss: 22.217537795193493

evaluating...

Confusion matrix, without normalization

[[ 238 45 11]

[1035 2701 85]

[ 139 25 678]]

Normalized confusion matrix

[[0.80952381 0.15306122 0.03741497]

[0.2708715 0.70688301 0.02224549]

[0.16508314 0.02969121 0.80522565]]





precision recall f1-score support

hate 0.17 0.81 0.28 294

offensive 0.97 0.71 0.82 3821

neither 0.88 0.81 0.84 842

accuracy 0.73 4957

macro avg 0.67 0.77 0.65 4957

weighted avg 0.91 0.73 0.79 4957

average eval\_loss: 0.8094781663328889, accuracy: 0.7296752067782933

training 5 epoch...

44.67226457595825 seconds used......

5 training finished! average train loss: 0.27617911316389054

total clf loss: 18.97973156347871 total adv loss: 14.123448439786443

evaluating...

Confusion matrix, without normalization

[[ 209 67 18]

[ 594 3120 107]

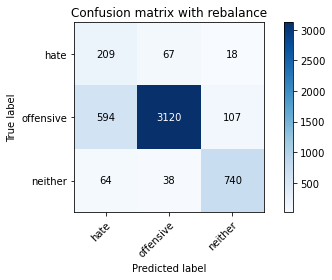
[ 64 38 740]]

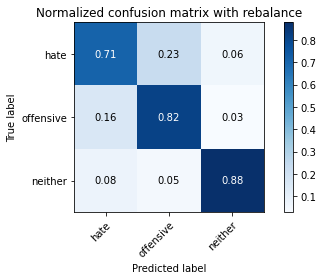
Normalized confusion matrix

[[0.71088435 0.22789116 0.06122449]

[0.15545669 0.81654017 0.02800314]

[0.0760095 0.04513064 0.87885986]]





precision recall f1-score support

hate 0.24 0.71 0.36 294

offensive 0.97 0.82 0.89 3821

neither 0.86 0.88 0.87 842

accuracy 0.82 4957

macro avg 0.69 0.80 0.70 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.6097190118875475, accuracy: 0.8208593907605406

training 6 epoch...

44.683659076690674 seconds used......

6 training finished! average train loss: 0.15344552126064984

total clf loss: 10.739970864029601 total adv loss: 7.749606605997542

evaluating...

Confusion matrix, without normalization

[[ 225 51 18]

[ 730 2964 127]

[ 79 23 740]]

Normalized confusion matrix

[[0.76530612 0.17346939 0.06122449]

[0.19104946 0.77571316 0.03323737]

[0.09382423 0.02731591 0.87885986]]





precision recall f1-score support

hate 0.22 0.77 0.34 294

offensive 0.98 0.78 0.86 3821

neither 0.84 0.88 0.86 842

accuracy 0.79 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.8416088662605973, accuracy: 0.7926165019164817

training 7 epoch...

44.71024799346924 seconds used......

7 training finished! average train loss: 0.06993598715682672

total clf loss: 5.48350159265101 total adv loss: 3.237776074269277

evaluating...

Confusion matrix, without normalization

[[ 216 60 18]

[ 641 3044 136]

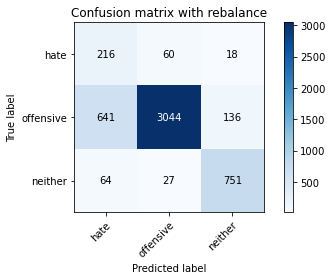
[ 64 27 751]]

Normalized confusion matrix

[[0.73469388 0.20408163 0.06122449]

[0.16775713 0.79665009 0.03559278]

[0.0760095 0.03206651 0.89192399]]





precision recall f1-score support

hate 0.23 0.73 0.36 294

offensive 0.97 0.80 0.88 3821

neither 0.83 0.89 0.86 842

accuracy 0.81 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.9244729324883308, accuracy: 0.8091587653822877

training 8 epoch...

44.660484790802 seconds used......

8 training finished! average train loss: 0.08494569080996148

total clf loss: 5.577308251929935 total adv loss: 4.474202484000671

evaluating...

Confusion matrix, without normalization

[[ 233 47 14]

[ 878 2809 134]

[ 87 23 732]]

Normalized confusion matrix

[[0.79251701 0.15986395 0.04761905]

[0.22978278 0.73514787 0.03506935]

[0.10332542 0.02731591 0.86935867]]





precision recall f1-score support

hate 0.19 0.79 0.31 294

offensive 0.98 0.74 0.84 3821

neither 0.83 0.87 0.85 842

accuracy 0.76 4957

macro avg 0.67 0.80 0.67 4957

weighted avg 0.90 0.76 0.81 4957

average eval\_loss: 1.1680564454291016, accuracy: 0.7613475892677022

training 9 epoch...

44.65716743469238 seconds used......

9 training finished! average train loss: 0.0641304742366522

total clf loss: 4.132456766034011 total adv loss: 3.41692715217323

evaluating...

Confusion matrix, without normalization

[[ 237 45 12]

[1077 2654 90]

[ 113 36 693]]

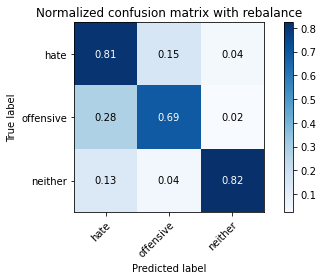
Normalized confusion matrix

[[0.80612245 0.15306122 0.04081633]

[0.28186339 0.69458257 0.02355404]

[0.13420428 0.04275534 0.82304038]]





precision recall f1-score support

hate 0.17 0.81 0.28 294

offensive 0.97 0.69 0.81 3821

neither 0.87 0.82 0.85 842

accuracy 0.72 4957

macro avg 0.67 0.77 0.64 4957

weighted avg 0.91 0.72 0.78 4957

average eval\_loss: 1.5240650464510244, accuracy: 0.7230179544079081

training 10 epoch...

44.72250175476074 seconds used......

10 training finished! average train loss: 0.03737798080117883

total clf loss: 2.397285284503596 total adv loss: 1.9971747264662554

evaluating...

Confusion matrix, without normalization

[[ 223 56 15]

[ 853 2853 115]

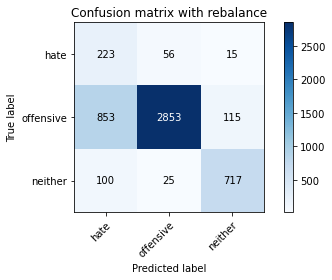
[ 100 25 717]]

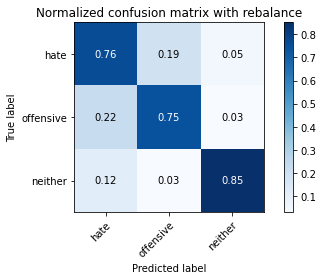
Normalized confusion matrix

[[0.7585034 0.19047619 0.05102041]

[0.22323999 0.74666318 0.03009683]

[0.11876485 0.02969121 0.85154394]]





precision recall f1-score support

hate 0.19 0.76 0.30 294

offensive 0.97 0.75 0.84 3821

neither 0.85 0.85 0.85 842

accuracy 0.77 4957

macro avg 0.67 0.79 0.67 4957

weighted avg 0.90 0.77 0.81 4957

average eval\_loss: 1.3550678466807209, accuracy: 0.7651805527536817

4 run

training 1 epoch...

44.700613498687744 seconds used......

1 training finished! average train loss: 1.51103255466411

total clf loss: 117.33057449758053 total adv loss: 70.52799696475267

evaluating...

Confusion matrix, without normalization

[[ 220 38 36]

[ 825 2862 134]

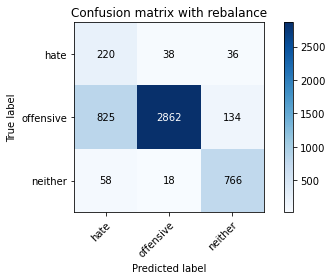
[ 58 18 766]]

Normalized confusion matrix

[[0.74829932 0.1292517 0.12244898]

[0.21591206 0.74901858 0.03506935]

[0.06888361 0.02137767 0.90973872]]





precision recall f1-score support

hate 0.20 0.75 0.31 294

offensive 0.98 0.75 0.85 3821

neither 0.82 0.91 0.86 842

accuracy 0.78 4957

macro avg 0.67 0.80 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 0.5366694758375806, accuracy: 0.7762759733709905

training 2 epoch...

44.636945724487305 seconds used......

2 training finished! average train loss: 1.0059643239654295

total clf loss: 72.11243622004986 total adv loss: 49.95373157411814

evaluating...

Confusion matrix, without normalization

[[ 220 54 20]

[ 636 3074 111]

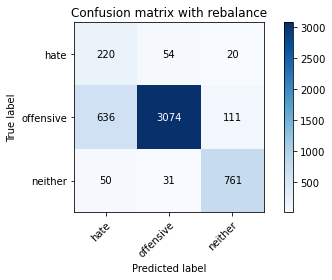
[ 50 31 761]]

Normalized confusion matrix

[[0.74829932 0.18367347 0.06802721]

[0.16644857 0.80450144 0.02904999]

[0.05938242 0.0368171 0.90380048]]





precision recall f1-score support

hate 0.24 0.75 0.37 294

offensive 0.97 0.80 0.88 3821

neither 0.85 0.90 0.88 842

accuracy 0.82 4957

macro avg 0.69 0.82 0.71 4957

weighted avg 0.91 0.82 0.85 4957

average eval\_loss: 0.43581814901723015, accuracy: 0.8180351018761347

training 3 epoch...

44.7592556476593 seconds used......

3 training finished! average train loss: 0.719050157749862

total clf loss: 51.26631799340248 total adv loss: 35.84562933817506

evaluating...

Confusion matrix, without normalization

[[ 187 71 36]

[ 637 2985 199]

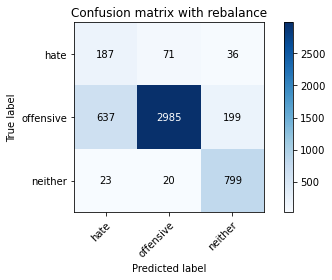
[ 23 20 799]]

Normalized confusion matrix

[[0.63605442 0.2414966 0.12244898]

[0.16671029 0.78120911 0.05208061]

[0.02731591 0.02375297 0.94893112]]





precision recall f1-score support

hate 0.22 0.64 0.33 294

offensive 0.97 0.78 0.87 3821

neither 0.77 0.95 0.85 842

accuracy 0.80 4957

macro avg 0.65 0.79 0.68 4957

weighted avg 0.89 0.80 0.83 4957

average eval\_loss: 0.4705524029991319, accuracy: 0.8010893685696994

training 4 epoch...

44.70370674133301 seconds used......

4 training finished! average train loss: 0.471000146817917

total clf loss: 34.167439782992005 total adv loss: 23.18679257086478

evaluating...

Confusion matrix, without normalization

[[ 231 43 20]

[ 804 2883 134]

[ 46 18 778]]

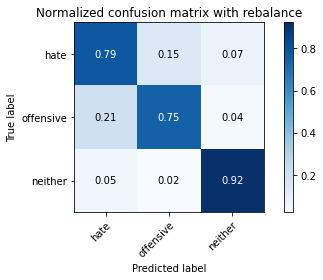
Normalized confusion matrix

[[0.78571429 0.1462585 0.06802721]

[0.21041612 0.75451452 0.03506935]

[0.05463183 0.02137767 0.9239905 ]]





precision recall f1-score support

hate 0.21 0.79 0.34 294

offensive 0.98 0.75 0.85 3821

neither 0.83 0.92 0.88 842

accuracy 0.79 4957

macro avg 0.68 0.82 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.6393861644150268, accuracy: 0.7851523098648376

training 5 epoch...

44.726261615753174 seconds used......

5 training finished! average train loss: 0.27195718235810074

total clf loss: 19.648487251251936 total adv loss: 13.428095424664207

evaluating...

Confusion matrix, without normalization

[[ 192 75 27]

[ 541 3139 141]

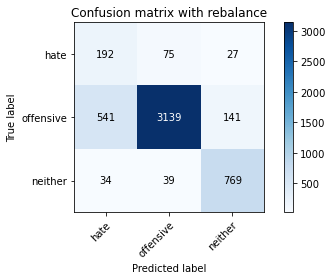
[ 34 39 769]]

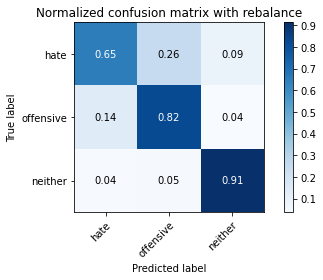
Normalized confusion matrix

[[0.65306122 0.25510204 0.09183673]

[0.14158597 0.82151269 0.03690133]

[0.04038005 0.04631829 0.91330166]]





precision recall f1-score support

hate 0.25 0.65 0.36 294

offensive 0.96 0.82 0.89 3821

neither 0.82 0.91 0.86 842

accuracy 0.83 4957

macro avg 0.68 0.80 0.70 4957

weighted avg 0.90 0.83 0.85 4957

average eval\_loss: 0.5681111453281295, accuracy: 0.8271131732902965

training 6 epoch...

44.73454189300537 seconds used......

6 training finished! average train loss: 0.16433719590807344

total clf loss: 11.728889795020223 total adv loss: 8.186385351786157

evaluating...

Confusion matrix, without normalization

[[ 212 57 25]

[ 659 2990 172]

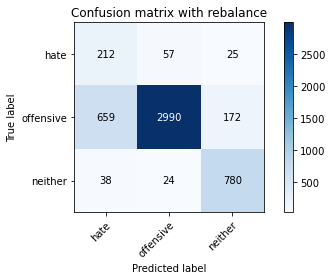
[ 38 24 780]]

Normalized confusion matrix

[[0.72108844 0.19387755 0.08503401]

[0.17246794 0.78251767 0.04501439]

[0.04513064 0.02850356 0.9263658 ]]





precision recall f1-score support

hate 0.23 0.72 0.35 294

offensive 0.97 0.78 0.87 3821

neither 0.80 0.93 0.86 842

accuracy 0.80 4957

macro avg 0.67 0.81 0.69 4957

weighted avg 0.90 0.80 0.84 4957

average eval\_loss: 0.7820313075602415, accuracy: 0.8033084526931612

training 7 epoch...

44.870676040649414 seconds used......

7 training finished! average train loss: 0.06213705421477142

total clf loss: 5.249412774690427 total adv loss: 2.688011739128342

evaluating...

Confusion matrix, without normalization

[[ 230 49 15]

[ 868 2830 123]

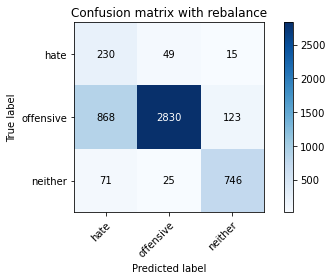
[ 71 25 746]]

Normalized confusion matrix

[[0.78231293 0.16666667 0.05102041]

[0.22716566 0.74064381 0.03219053]

[0.08432304 0.02969121 0.88598575]]





precision recall f1-score support

hate 0.20 0.78 0.31 294

offensive 0.97 0.74 0.84 3821

neither 0.84 0.89 0.86 842

accuracy 0.77 4957

macro avg 0.67 0.80 0.67 4957

weighted avg 0.91 0.77 0.81 4957

average eval\_loss: 1.067307960636343, accuracy: 0.7678031067177729

training 8 epoch...

44.78736066818237 seconds used......

8 training finished! average train loss: 0.07977809918750274

total clf loss: 5.305063887266442 total adv loss: 4.168495528087988

evaluating...

Confusion matrix, without normalization

[[ 205 68 21]

[ 575 3107 139]

[ 46 30 766]]

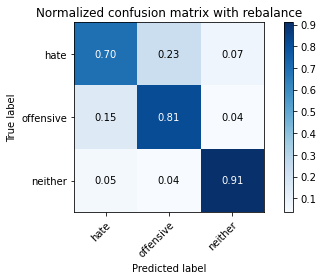
Normalized confusion matrix

[[0.69727891 0.23129252 0.07142857]

[0.15048417 0.81313792 0.03637791]

[0.05463183 0.03562945 0.90973872]]





precision recall f1-score support

hate 0.25 0.70 0.37 294

offensive 0.97 0.81 0.88 3821

neither 0.83 0.91 0.87 842

accuracy 0.82 4957

macro avg 0.68 0.81 0.71 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.891072110475732, accuracy: 0.822675005043373

training 9 epoch...

44.906733751297 seconds used......

9 training finished! average train loss: 0.06491621715092911

total clf loss: 4.464715457172133 total adv loss: 3.3179788457673567

evaluating...

Confusion matrix, without normalization

[[ 210 65 19]

[ 600 3094 127]

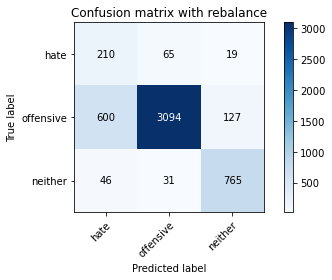
[ 46 31 765]]

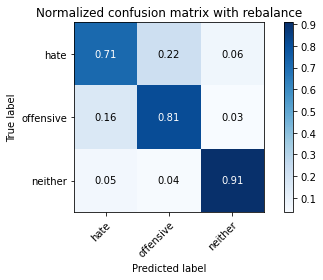
Normalized confusion matrix

[[0.71428571 0.22108844 0.06462585]

[0.15702696 0.80973567 0.03323737]

[0.05463183 0.0368171 0.90855107]]





precision recall f1-score support

hate 0.25 0.71 0.37 294

offensive 0.97 0.81 0.88 3821

neither 0.84 0.91 0.87 842

accuracy 0.82 4957

macro avg 0.68 0.81 0.71 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.9193844714524946, accuracy: 0.8208593907605406

training 10 epoch...

44.81362724304199 seconds used......

10 training finished! average train loss: 0.04879697176358221

total clf loss: 3.1209208501677494 total adv loss: 2.611680651158167

evaluating...

Confusion matrix, without normalization

[[ 237 38 19]

[ 943 2740 138]

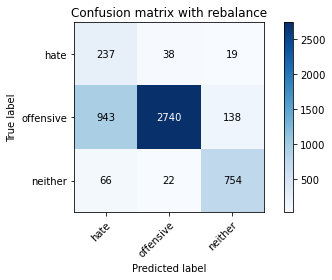
[ 66 22 754]]

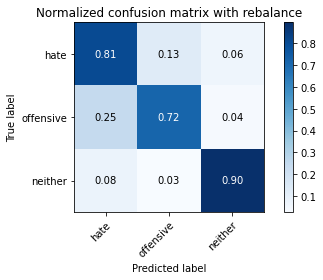
Normalized confusion matrix

[[0.80612245 0.1292517 0.06462585]

[0.24679403 0.71708977 0.0361162 ]

[0.0783848 0.02612827 0.89548694]]





precision recall f1-score support

hate 0.19 0.81 0.31 294

offensive 0.98 0.72 0.83 3821

neither 0.83 0.90 0.86 842

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: 1.4518167765570744, accuracy: 0.7526729876941699

5 run

training 1 epoch...

44.848134994506836 seconds used......

1 training finished! average train loss: 1.5433936352618256

total clf loss: 120.05385166406631 total adv loss: 71.9332303032279

evaluating...

Confusion matrix, without normalization

[[ 193 57 44]

[ 579 3112 130]

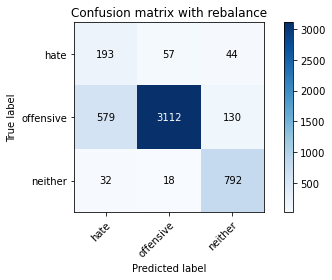
[ 32 18 792]]

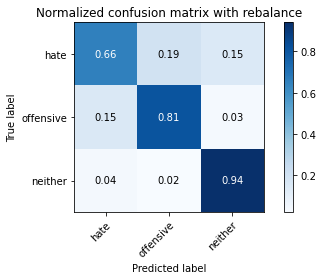
Normalized confusion matrix

[[0.65646259 0.19387755 0.14965986]

[0.15153101 0.81444648 0.03402251]

[0.03800475 0.02137767 0.94061758]]





precision recall f1-score support

hate 0.24 0.66 0.35 294

offensive 0.98 0.81 0.89 3821

neither 0.82 0.94 0.88 842

accuracy 0.83 4957

macro avg 0.68 0.80 0.71 4957

weighted avg 0.91 0.83 0.85 4957

average eval\_loss: 0.3948114790743397, accuracy: 0.8265079685293524

training 2 epoch...

44.76909041404724 seconds used......

2 training finished! average train loss: 1.002082160864657

total clf loss: 71.85496842861176 total adv loss: 49.750540643930435

evaluating...

Confusion matrix, without normalization

[[ 219 53 22]

[ 680 2979 162]

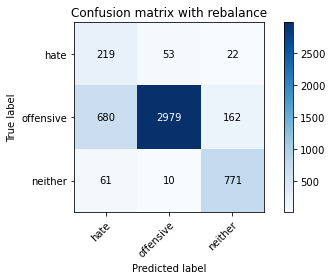
[ 61 10 771]]

Normalized confusion matrix

[[0.74489796 0.18027211 0.07482993]

[0.17796388 0.77963884 0.04239728]

[0.07244656 0.01187648 0.91567696]]





precision recall f1-score support

hate 0.23 0.74 0.35 294

offensive 0.98 0.78 0.87 3821

neither 0.81 0.92 0.86 842

accuracy 0.80 4957

macro avg 0.67 0.81 0.69 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.5058080108175355, accuracy: 0.80068589872907

training 3 epoch...

44.71873116493225 seconds used......

3 training finished! average train loss: 0.7143419476454718

total clf loss: 50.59846551716328 total adv loss: 35.77700366731733

evaluating...

Confusion matrix, without normalization

[[ 242 35 17]

[ 994 2690 137]

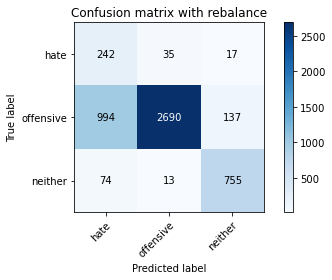
[ 74 13 755]]

Normalized confusion matrix

[[0.82312925 0.11904762 0.05782313]

[0.26014132 0.70400419 0.03585449]

[0.08788599 0.01543943 0.89667458]]





precision recall f1-score support

hate 0.18 0.82 0.30 294

offensive 0.98 0.70 0.82 3821

neither 0.83 0.90 0.86 842

accuracy 0.74 4957

macro avg 0.67 0.81 0.66 4957

weighted avg 0.91 0.74 0.80 4957

average eval\_loss: 0.7155998662684954, accuracy: 0.7437966512003228

training 4 epoch...

44.65500450134277 seconds used......

4 training finished! average train loss: 0.39434289728558203

total clf loss: 28.183224748820066 total adv loss: 19.62470531417057

evaluating...

Confusion matrix, without normalization

[[ 215 66 13]

[ 636 3104 81]

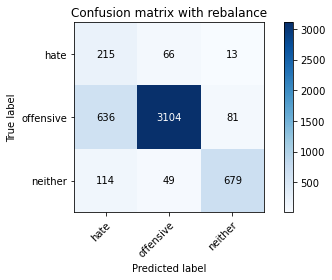
[ 114 49 679]]

Normalized confusion matrix

[[0.73129252 0.2244898 0.04421769]

[0.16644857 0.81235279 0.02119864]

[0.13539192 0.05819477 0.8064133 ]]





precision recall f1-score support

hate 0.22 0.73 0.34 294

offensive 0.96 0.81 0.88 3821

neither 0.88 0.81 0.84 842

accuracy 0.81 4957

macro avg 0.69 0.78 0.69 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 0.5629030056297779, accuracy: 0.8065362114181965

training 5 epoch...

44.68114876747131 seconds used......

5 training finished! average train loss: 0.21812442830649384

total clf loss: 15.07149781472981 total adv loss: 11.113889682572335

evaluating...

Confusion matrix, without normalization

[[ 211 62 21]

[ 695 2995 131]

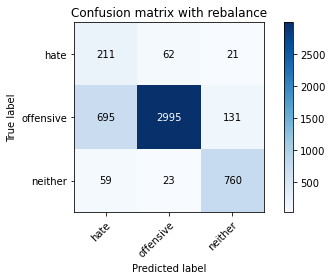
[ 59 23 760]]

Normalized confusion matrix

[[0.71768707 0.21088435 0.07142857]

[0.18188956 0.78382622 0.03428422]

[0.07007126 0.02731591 0.90261283]]





precision recall f1-score support

hate 0.22 0.72 0.34 294

offensive 0.97 0.78 0.87 3821

neither 0.83 0.90 0.87 842

accuracy 0.80 4957

macro avg 0.67 0.80 0.69 4957

weighted avg 0.90 0.80 0.84 4957

average eval\_loss: 0.7726948740291258, accuracy: 0.8000806939681259

training 6 epoch...

44.63548278808594 seconds used......

6 training finished! average train loss: 0.11688464167337233

total clf loss: 9.055438549024984 total adv loss: 5.465917538931535

evaluating...

Confusion matrix, without normalization

[[ 234 42 18]

[ 945 2771 105]

[ 80 28 734]]

Normalized confusion matrix

[[0.79591837 0.14285714 0.06122449]

[0.24731746 0.72520283 0.02747972]

[0.09501188 0.03325416 0.87173397]]





precision recall f1-score support

hate 0.19 0.80 0.30 294

offensive 0.98 0.73 0.83 3821

neither 0.86 0.87 0.86 842

accuracy 0.75 4957

macro avg 0.67 0.80 0.67 4957

weighted avg 0.91 0.75 0.81 4957

average eval\_loss: 1.1536954166217437, accuracy: 0.7542868670566875

training 7 epoch...

44.66652035713196 seconds used......

7 training finished! average train loss: 0.11764607591723358

total clf loss: 7.654763100203127 total adv loss: 6.231357937052962

evaluating...

Confusion matrix, without normalization

[[ 200 70 24]

[ 502 3124 195]

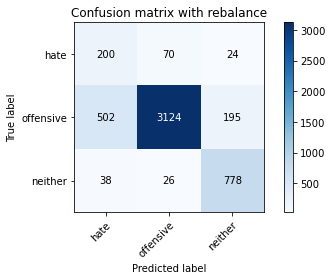
[ 38 26 778]]

Normalized confusion matrix

[[0.68027211 0.23809524 0.08163265]

[0.13137922 0.81758702 0.05103376]

[0.04513064 0.03087886 0.9239905 ]]





precision recall f1-score support

hate 0.27 0.68 0.39 294

offensive 0.97 0.82 0.89 3821

neither 0.78 0.92 0.85 842

accuracy 0.83 4957

macro avg 0.67 0.81 0.71 4957

weighted avg 0.90 0.83 0.85 4957

average eval\_loss: 0.7801485409811815, accuracy: 0.827516643130926

training 8 epoch...

44.68441653251648 seconds used......

8 training finished! average train loss: 0.08003987320045237

total clf loss: 5.728136261983309 total adv loss: 3.9793410498532467

evaluating...

Confusion matrix, without normalization

[[ 218 55 21]

[ 738 2919 164]

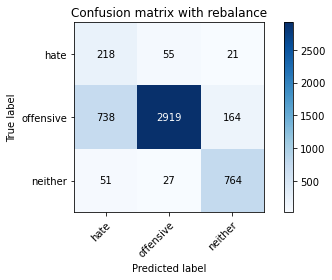
[ 51 27 764]]

Normalized confusion matrix

[[0.7414966 0.18707483 0.07142857]

[0.19314316 0.76393614 0.0429207 ]

[0.06057007 0.03206651 0.90736342]]





precision recall f1-score support

hate 0.22 0.74 0.34 294

offensive 0.97 0.76 0.86 3821

neither 0.81 0.91 0.85 842

accuracy 0.79 4957

macro avg 0.66 0.80 0.68 4957

weighted avg 0.90 0.79 0.82 4957

average eval\_loss: 0.9979478096703984, accuracy: 0.7869679241476699

training 9 epoch...

44.636746644973755 seconds used......

9 training finished! average train loss: 0.047326444066810414

total clf loss: 3.2586668437579647 total adv loss: 2.417077533871634

evaluating...

Confusion matrix, without normalization

[[ 236 46 12]

[ 839 2885 97]

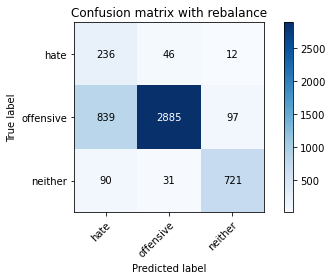
[ 90 31 721]]

Normalized confusion matrix

[[0.80272109 0.15646259 0.04081633]

[0.21957603 0.75503795 0.02538602]

[0.10688836 0.0368171 0.85629454]]





precision recall f1-score support

hate 0.20 0.80 0.32 294

offensive 0.97 0.76 0.85 3821

neither 0.87 0.86 0.86 842

accuracy 0.78 4957

macro avg 0.68 0.80 0.68 4957

weighted avg 0.91 0.78 0.82 4957

average eval\_loss: 1.2206310478223097, accuracy: 0.7750655638491023

training 10 epoch...

44.584667444229126 seconds used......

10 training finished! average train loss: 0.015728572645797856

total clf loss: 1.5127545660361648 total adv loss: 0.5884156710308162

evaluating...

Confusion matrix, without normalization

[[ 232 46 16]

[ 824 2844 153]

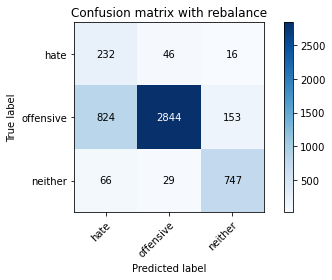
[ 66 29 747]]

Normalized confusion matrix

[[0.78911565 0.15646259 0.05442177]

[0.21565035 0.74430777 0.04004187]

[0.0783848 0.03444181 0.8871734 ]]





precision recall f1-score support

hate 0.21 0.79 0.33 294

offensive 0.97 0.74 0.84 3821

neither 0.82 0.89 0.85 842

accuracy 0.77 4957

macro avg 0.67 0.81 0.67 4957

weighted avg 0.90 0.77 0.81 4957

average eval\_loss: 1.2798772915369725, accuracy: 0.7712326003631229