hate as offensive

alpha=0.05

input: last layer

training 1 epoch...

50.02083969116211 seconds used......

1 training finished! average train loss: 0.6424809482536817

total clf loss: 112.57467399537563 total adv loss: 54.208636444061995

evaluating...

Confusion matrix, without normalization

[[ 259 26 9]

[1467 2311 43]

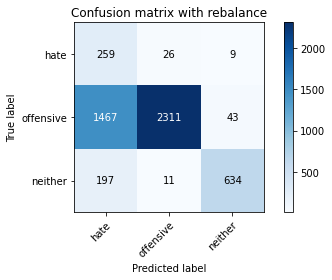
[ 197 11 634]]

Normalized confusion matrix

[[0.88095238 0.08843537 0.03061224]

[0.38393091 0.60481549 0.0112536 ]

[0.23396675 0.01306413 0.75296912]]





precision recall f1-score support

hate 0.13 0.88 0.23 294

offensive 0.98 0.60 0.75 3821

neither 0.92 0.75 0.83 842

accuracy 0.65 4957

macro avg 0.68 0.75 0.60 4957

weighted avg 0.92 0.65 0.73 4957

average eval\_loss: 0.8290862605696724, accuracy: 0.6463586846883196

training 2 epoch...

49.99864387512207 seconds used......

2 training finished! average train loss: 0.40222456179864224

total clf loss: 70.1465257331729 total adv loss: 27.322513956576586

evaluating...

Confusion matrix, without normalization

[[ 235 46 13]

[ 723 3003 95]

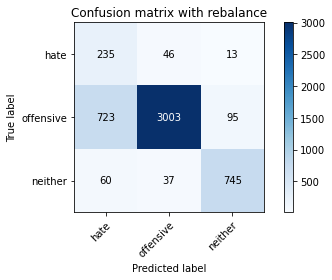
[ 60 37 745]]

Normalized confusion matrix

[[0.79931973 0.15646259 0.04421769]

[0.18921748 0.78591992 0.0248626 ]

[0.07125891 0.04394299 0.8847981 ]]





precision recall f1-score support

hate 0.23 0.80 0.36 294

offensive 0.97 0.79 0.87 3821

neither 0.87 0.88 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: 0.4376287110630543, accuracy: 0.8035101876134759

training 3 epoch...

49.947346448898315 seconds used......

3 training finished! average train loss: 0.26419503113244014

total clf loss: 45.87072954699397 total adv loss: 13.867581768427044

evaluating...

Confusion matrix, without normalization

[[ 235 47 12]

[ 776 2960 85]

[ 102 39 701]]

Normalized confusion matrix

[[0.79931973 0.15986395 0.04081633]

[0.2030882 0.77466632 0.02224549]

[0.12114014 0.04631829 0.83254157]]





precision recall f1-score support

hate 0.21 0.80 0.33 294

offensive 0.97 0.77 0.86 3821

neither 0.88 0.83 0.85 842

accuracy 0.79 4957

macro avg 0.69 0.80 0.68 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.49200361879963067, accuracy: 0.7859592495460964

training 4 epoch...

49.97953748703003 seconds used......

4 training finished! average train loss: 0.1649162262427615

total clf loss: 28.43983045592904 total adv loss: 4.7831136585446075

evaluating...

Confusion matrix, without normalization

[[ 236 47 11]

[ 946 2806 69]

[ 90 60 692]]

Normalized confusion matrix

[[0.80272109 0.15986395 0.03741497]

[0.24757917 0.73436273 0.0180581 ]

[0.10688836 0.07125891 0.82185273]]





precision recall f1-score support

hate 0.19 0.80 0.30 294

offensive 0.96 0.73 0.83 3821

neither 0.90 0.82 0.86 842

accuracy 0.75 4957

macro avg 0.68 0.79 0.66 4957

weighted avg 0.91 0.75 0.81 4957

average eval\_loss: 0.6565529193068224, accuracy: 0.753278192455114

training 5 epoch...

49.973833084106445 seconds used......

5 training finished! average train loss: 0.09275157375217617

total clf loss: 16.095199247356504 total adv loss: 4.693603671999881

evaluating...

Confusion matrix, without normalization

[[ 208 66 20]

[ 567 3106 148]

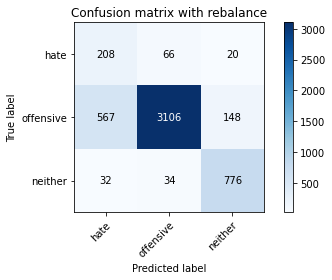
[ 32 34 776]]

Normalized confusion matrix

[[0.70748299 0.2244898 0.06802721]

[0.14839047 0.81287621 0.03873332]

[0.03800475 0.04038005 0.9216152 ]]





precision recall f1-score support

hate 0.26 0.71 0.38 294

offensive 0.97 0.81 0.88 3821

neither 0.82 0.92 0.87 842

accuracy 0.83 4957

macro avg 0.68 0.81 0.71 4957

weighted avg 0.90 0.83 0.85 4957

average eval\_loss: 0.5882752580012405, accuracy: 0.8250958240871494

training 6 epoch...

50.047579288482666 seconds used......

6 training finished! average train loss: 0.040833273219020916

total clf loss: 7.109957718057558 total adv loss: 2.549360265111318

evaluating...

Confusion matrix, without normalization

[[ 228 53 13]

[ 813 2912 96]

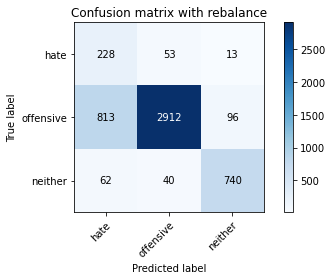
[ 62 40 740]]

Normalized confusion matrix

[[0.7755102 0.18027211 0.04421769]

[0.21277153 0.76210416 0.02512431]

[0.0736342 0.04750594 0.87885986]]





precision recall f1-score support

hate 0.21 0.78 0.33 294

offensive 0.97 0.76 0.85 3821

neither 0.87 0.88 0.88 842

accuracy 0.78 4957

macro avg 0.68 0.81 0.68 4957

weighted avg 0.91 0.78 0.83 4957

average eval\_loss: 0.9344079353426012, accuracy: 0.7827314908210611

training 7 epoch...

49.99364447593689 seconds used......

7 training finished! average train loss: 0.023374354558021962

total clf loss: 4.008629450167064 total adv loss: 0.23229657861520536

evaluating...

Confusion matrix, without normalization

[[ 225 52 17]

[ 696 3014 111]

[ 47 36 759]]

Normalized confusion matrix

[[0.76530612 0.17687075 0.05782313]

[0.18215127 0.78879874 0.02904999]

[0.05581948 0.04275534 0.90142518]]





precision recall f1-score support

hate 0.23 0.77 0.36 294

offensive 0.97 0.79 0.87 3821

neither 0.86 0.90 0.88 842

accuracy 0.81 4957

macro avg 0.69 0.82 0.70 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 0.9068052668815418, accuracy: 0.8065362114181965

training 8 epoch...

49.92884874343872 seconds used......

8 training finished! average train loss: 0.03126983741728578

total clf loss: 5.481041379622184 total adv loss: 2.6779834926419426

evaluating...

Confusion matrix, without normalization

[[ 225 50 19]

[ 841 2848 132]

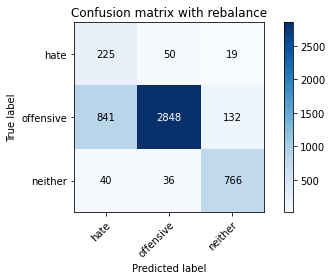
[ 40 36 766]]

Normalized confusion matrix

[[0.76530612 0.17006803 0.06462585]

[0.22009945 0.74535462 0.03454593]

[0.04750594 0.04275534 0.90973872]]





precision recall f1-score support

hate 0.20 0.77 0.32 294

offensive 0.97 0.75 0.84 3821

neither 0.84 0.91 0.87 842

accuracy 0.77 4957

macro avg 0.67 0.81 0.68 4957

weighted avg 0.90 0.77 0.82 4957

average eval\_loss: 1.0814429924583002, accuracy: 0.7744603590881581

training 9 epoch...

49.98467993736267 seconds used......

9 training finished! average train loss: 0.01260934700428274

total clf loss: 2.2006149790249765 total adv loss: 0.8883330075550475

evaluating...

Confusion matrix, without normalization

[[ 225 56 13]

[ 728 3010 83]

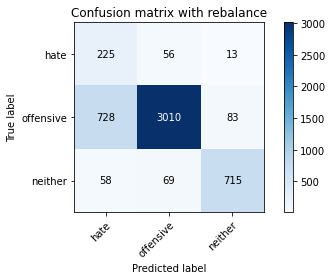
[ 58 69 715]]

Normalized confusion matrix

[[0.76530612 0.19047619 0.04421769]

[0.19052604 0.7877519 0.02172206]

[0.06888361 0.08194774 0.84916865]]





precision recall f1-score support

hate 0.22 0.77 0.34 294

offensive 0.96 0.79 0.87 3821

neither 0.88 0.85 0.87 842

accuracy 0.80 4957

macro avg 0.69 0.80 0.69 4957

weighted avg 0.90 0.80 0.83 4957

average eval\_loss: 0.9817852414778299, accuracy: 0.7968529352430905

training 10 epoch...

50.041173458099365 seconds used......

10 training finished! average train loss: 0.013096277369016085

total clf loss: 2.2533922170987353 total adv loss: 0.2785757349411142

evaluating...

Confusion matrix, without normalization

[[ 222 55 17]

[ 832 2864 125]

[ 46 41 755]]

Normalized confusion matrix

[[0.75510204 0.18707483 0.05782313]

[0.21774405 0.749542 0.03271395]

[0.05463183 0.04869359 0.89667458]]





precision recall f1-score support

hate 0.20 0.76 0.32 294

offensive 0.97 0.75 0.84 3821

neither 0.84 0.90 0.87 842

accuracy 0.77 4957

macro avg 0.67 0.80 0.68 4957

weighted avg 0.90 0.77 0.82 4957

average eval\_loss: 1.159336865121018, accuracy: 0.7748638289287876

2 run...

training 1 epoch...

49.97576594352722 seconds used......

1 training finished! average train loss: 0.669962907220885

total clf loss: 117.43885217607021 total adv loss: 57.50389663502574

evaluating...

Confusion matrix, without normalization

[[ 224 50 20]

[ 772 2940 109]

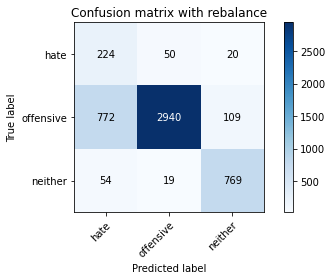
[ 54 19 769]]

Normalized confusion matrix

[[0.76190476 0.17006803 0.06802721]

[0.20204135 0.76943209 0.02852656]

[0.06413302 0.02256532 0.91330166]]





precision recall f1-score support

hate 0.21 0.76 0.33 294

offensive 0.98 0.77 0.86 3821

neither 0.86 0.91 0.88 842

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: 0.46377161013022544, accuracy: 0.7934234415977406

training 2 epoch...

49.9896457195282 seconds used......

2 training finished! average train loss: 0.40467545790979037

total clf loss: 70.60995782911777 total adv loss: 28.20909099560231

evaluating...

Confusion matrix, without normalization

[[ 214 61 19]

[ 606 3104 111]

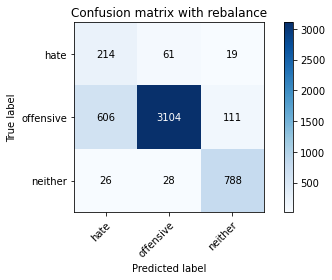
[ 26 28 788]]

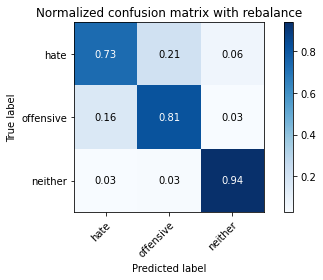
Normalized confusion matrix

[[0.72789116 0.20748299 0.06462585]

[0.15859723 0.81235279 0.02904999]

[0.03087886 0.03325416 0.93586698]]





precision recall f1-score support

hate 0.25 0.73 0.38 294

offensive 0.97 0.81 0.89 3821

neither 0.86 0.94 0.90 842

accuracy 0.83 4957

macro avg 0.69 0.83 0.72 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.3938605437475827, accuracy: 0.8283235828121848

training 3 epoch...

49.91573739051819 seconds used......

3 training finished! average train loss: 0.2641999869885152

total clf loss: 45.86843268945813 total adv loss: 13.804700461565517

evaluating...

Confusion matrix, without normalization

[[ 234 47 13]

[ 779 2949 93]

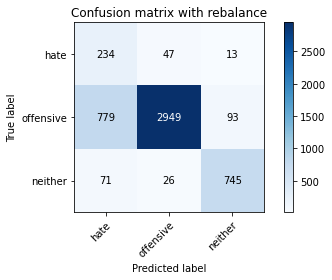
[ 71 26 745]]

Normalized confusion matrix

[[0.79591837 0.15986395 0.04421769]

[0.20387333 0.77178749 0.02433918]

[0.08432304 0.03087886 0.8847981 ]]





precision recall f1-score support

hate 0.22 0.80 0.34 294

offensive 0.98 0.77 0.86 3821

neither 0.88 0.88 0.88 842

accuracy 0.79 4957

macro avg 0.69 0.82 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.5484969191733868, accuracy: 0.7924147669961671

training 4 epoch...

49.92683219909668 seconds used......

4 training finished! average train loss: 0.1640643777750563

total clf loss: 28.43354799039662 total adv loss: 7.57078656257363

evaluating...

Confusion matrix, without normalization

[[ 229 51 14]

[ 746 2981 94]

[ 64 42 736]]

Normalized confusion matrix

[[0.77891156 0.17346939 0.04761905]

[0.19523685 0.78016226 0.02460089]

[0.0760095 0.04988124 0.87410926]]





precision recall f1-score support

hate 0.22 0.78 0.34 294

offensive 0.97 0.78 0.86 3821

neither 0.87 0.87 0.87 842

accuracy 0.80 4957

macro avg 0.69 0.81 0.69 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.5469296120647942, accuracy: 0.7960459955618318

training 5 epoch...

50.02808618545532 seconds used......

5 training finished! average train loss: 0.08480963075522617

total clf loss: 14.629430517554283 total adv loss: 2.539673534571193

evaluating...

Confusion matrix, without normalization

[[ 221 57 16]

[ 745 2959 117]

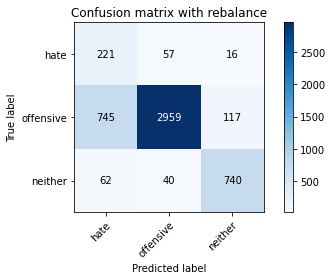
[ 62 40 740]]

Normalized confusion matrix

[[0.75170068 0.19387755 0.05442177]

[0.19497514 0.77440461 0.03062026]

[0.0736342 0.04750594 0.87885986]]





precision recall f1-score support

hate 0.21 0.75 0.33 294

offensive 0.97 0.77 0.86 3821

neither 0.85 0.88 0.86 842

accuracy 0.79 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.90 0.79 0.83 4957

average eval\_loss: 0.7181877451648395, accuracy: 0.7908008876336494

training 6 epoch...

50.00326728820801 seconds used......

6 training finished! average train loss: 0.035121826676935045

total clf loss: 6.078846111893654 total adv loss: 1.4602749394252896

evaluating...

Confusion matrix, without normalization

[[ 217 59 18]

[ 709 2999 113]

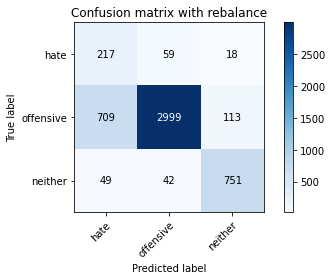
[ 49 42 751]]

Normalized confusion matrix

[[0.73809524 0.20068027 0.06122449]

[0.18555352 0.78487307 0.02957341]

[0.05819477 0.04988124 0.89192399]]





precision recall f1-score support

hate 0.22 0.74 0.34 294

offensive 0.97 0.78 0.87 3821

neither 0.85 0.89 0.87 842

accuracy 0.80 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.90 0.80 0.84 4957

average eval\_loss: 0.8235431431094185, accuracy: 0.8002824288884406

training 7 epoch...

49.949302434921265 seconds used......

7 training finished! average train loss: 0.02097381601533553

total clf loss: 3.672178333508782 total adv loss: 1.713116012266255

evaluating...

Confusion matrix, without normalization

[[ 203 70 21]

[ 593 3094 134]

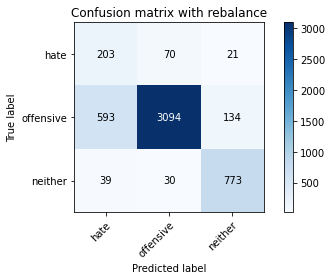
[ 39 30 773]]

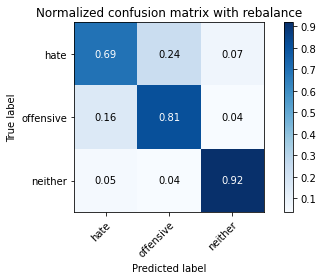
Normalized confusion matrix

[[0.69047619 0.23809524 0.07142857]

[0.15519498 0.80973567 0.03506935]

[0.04631829 0.03562945 0.91805226]]





precision recall f1-score support

hate 0.24 0.69 0.36 294

offensive 0.97 0.81 0.88 3821

neither 0.83 0.92 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: 0.8617731537331917, accuracy: 0.8210611256808553

training 8 epoch...

49.98630452156067 seconds used......

8 training finished! average train loss: 0.01804427055309102

total clf loss: 3.1413073924777564 total adv loss: 1.1147432835350628

evaluating...

Confusion matrix, without normalization

[[ 218 61 15]

[ 704 3033 84]

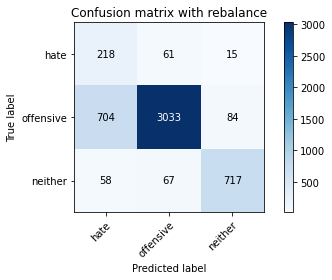
[ 58 67 717]]

Normalized confusion matrix

[[0.7414966 0.20748299 0.05102041]

[0.18424496 0.79377126 0.02198377]

[0.06888361 0.07957245 0.85154394]]





precision recall f1-score support

hate 0.22 0.74 0.34 294

offensive 0.96 0.79 0.87 3821

neither 0.88 0.85 0.86 842

accuracy 0.80 4957

macro avg 0.69 0.80 0.69 4957

weighted avg 0.90 0.80 0.84 4957

average eval\_loss: 1.0050957492766752, accuracy: 0.8004841638087553

training 9 epoch...

49.9972779750824 seconds used......

9 training finished! average train loss: 0.025404950312634234

total clf loss: 4.38852499355562 total adv loss: 0.8855698248153203

evaluating...

Confusion matrix, without normalization

[[ 234 43 17]

[ 928 2804 89]

[ 63 43 736]]

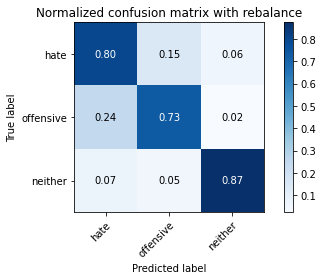
Normalized confusion matrix

[[0.79591837 0.1462585 0.05782313]

[0.24286836 0.73383931 0.02329233]

[0.07482185 0.05106888 0.87410926]]





precision recall f1-score support

hate 0.19 0.80 0.31 294

offensive 0.97 0.73 0.84 3821

neither 0.87 0.87 0.87 842

accuracy 0.76 4957

macro avg 0.68 0.80 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 1.1905861477998476, accuracy: 0.7613475892677022

training 10 epoch...

50.03523278236389 seconds used......

10 training finished! average train loss: 0.006533147399218437

total clf loss: 1.1671870675927494 total adv loss: 1.0003770971816266

evaluating...

Confusion matrix, without normalization

[[ 230 46 18]

[ 786 2913 122]

[ 60 32 750]]

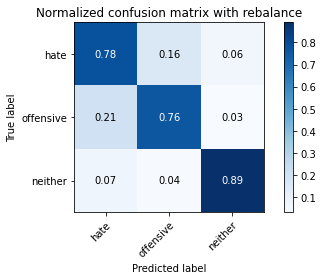
Normalized confusion matrix

[[0.78231293 0.15646259 0.06122449]

[0.20570531 0.76236587 0.03192881]

[0.07125891 0.03800475 0.89073634]]





precision recall f1-score support

hate 0.21 0.78 0.34 294

offensive 0.97 0.76 0.86 3821

neither 0.84 0.89 0.87 842

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: 1.228269805160967, accuracy: 0.7853540447851524

3 run...

training 1 epoch...

49.996840715408325 seconds used......

1 training finished! average train loss: 0.6382267516084582

total clf loss: 111.84073860943317 total adv loss: 54.07928014732897

evaluating...

Confusion matrix, without normalization

[[ 230 44 20]

[ 807 2903 111]

[ 59 23 760]]

Normalized confusion matrix

[[0.78231293 0.14965986 0.06802721]

[0.21120126 0.75974876 0.02904999]

[0.07007126 0.02731591 0.90261283]]





precision recall f1-score support

hate 0.21 0.78 0.33 294

offensive 0.98 0.76 0.85 3821

neither 0.85 0.90 0.88 842

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: 0.5045120190348356, accuracy: 0.7853540447851524

training 2 epoch...

50.06937861442566 seconds used......

2 training finished! average train loss: 0.387057235453561

total clf loss: 67.58080498874187 total adv loss: 27.880358899012208

evaluating...

Confusion matrix, without normalization

[[ 207 67 20]

[ 425 3295 101]

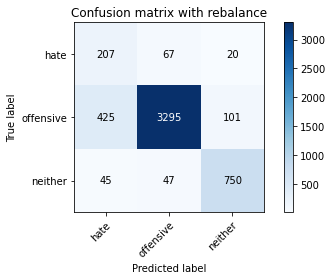
[ 45 47 750]]

Normalized confusion matrix

[[0.70408163 0.22789116 0.06802721]

[0.11122743 0.8623397 0.02643287]

[0.05344418 0.05581948 0.89073634]]





precision recall f1-score support

hate 0.31 0.70 0.43 294

offensive 0.97 0.86 0.91 3821

neither 0.86 0.89 0.88 842

accuracy 0.86 4957

macro avg 0.71 0.82 0.74 4957

weighted avg 0.91 0.86 0.88 4957

average eval\_loss: 0.38585883997861414, accuracy: 0.8577768811781319

training 3 epoch...

49.993366956710815 seconds used......

3 training finished! average train loss: 0.2613535365906724

total clf loss: 45.389069605618715 total adv loss: 13.952295824885368

evaluating...

Confusion matrix, without normalization

[[ 206 71 17]

[ 499 3206 116]

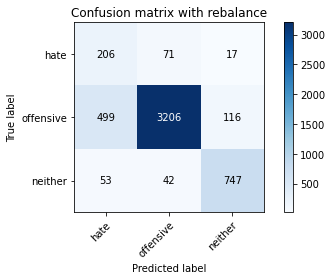
[ 53 42 747]]

Normalized confusion matrix

[[0.70068027 0.2414966 0.05782313]

[0.13059409 0.83904737 0.03035854]

[0.06294537 0.04988124 0.8871734 ]]





precision recall f1-score support

hate 0.27 0.70 0.39 294

offensive 0.97 0.84 0.90 3821

neither 0.85 0.89 0.87 842

accuracy 0.84 4957

macro avg 0.70 0.81 0.72 4957

weighted avg 0.90 0.84 0.86 4957

average eval\_loss: 0.42722110379667533, accuracy: 0.8390155335888643

training 4 epoch...

49.983506202697754 seconds used......

4 training finished! average train loss: 0.1479819739422603

total clf loss: 25.532174173742533 total adv loss: 4.545133967185393

evaluating...

Confusion matrix, without normalization

[[ 217 57 20]

[ 656 3044 121]

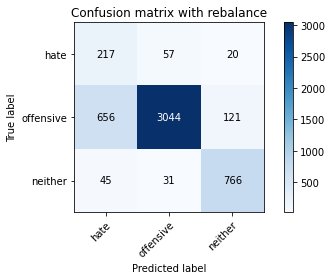
[ 45 31 766]]

Normalized confusion matrix

[[0.73809524 0.19387755 0.06802721]

[0.17168281 0.79665009 0.0316671 ]

[0.05344418 0.0368171 0.90973872]]





precision recall f1-score support

hate 0.24 0.74 0.36 294

offensive 0.97 0.80 0.88 3821

neither 0.84 0.91 0.88 842

accuracy 0.81 4957

macro avg 0.68 0.81 0.70 4957

weighted avg 0.91 0.81 0.84 4957

average eval\_loss: 0.5588606941183248, accuracy: 0.8123865241073229

training 5 epoch...

50.03054118156433 seconds used......

5 training finished! average train loss: 0.06855493952787062

total clf loss: 11.825064088217914 total adv loss: 2.0433895024616504

evaluating...

Confusion matrix, without normalization

[[ 219 59 16]

[ 739 2973 109]

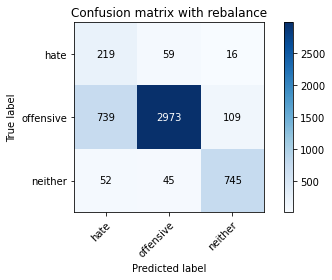
[ 52 45 745]]

Normalized confusion matrix

[[0.74489796 0.20068027 0.05442177]

[0.19340487 0.77806857 0.02852656]

[0.06175772 0.05344418 0.8847981 ]]





precision recall f1-score support

hate 0.22 0.74 0.34 294

offensive 0.97 0.78 0.86 3821

neither 0.86 0.88 0.87 842

accuracy 0.79 4957

macro avg 0.68 0.80 0.69 4957

weighted avg 0.90 0.79 0.83 4957

average eval\_loss: 0.727050062299015, accuracy: 0.7942303812789994

training 6 epoch...

49.951338052749634 seconds used......

6 training finished! average train loss: 0.04456356919386931

total clf loss: 7.738795499666594 total adv loss: 2.3685034957889

evaluating...

Confusion matrix, without normalization

[[ 205 71 18]

[ 664 3058 99]

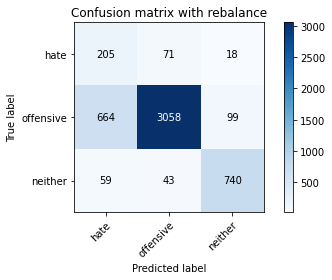
[ 59 43 740]]

Normalized confusion matrix

[[0.69727891 0.2414966 0.06122449]

[0.1737765 0.80031405 0.02590945]

[0.07007126 0.05106888 0.87885986]]





precision recall f1-score support

hate 0.22 0.70 0.34 294

offensive 0.96 0.80 0.87 3821

neither 0.86 0.88 0.87 842

accuracy 0.81 4957

macro avg 0.68 0.79 0.69 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.7791605775973068, accuracy: 0.80754488601977

training 7 epoch...

49.98791003227234 seconds used......

7 training finished! average train loss: 0.029511240912105736

total clf loss: 5.139884103322402 total adv loss: 1.8692383559973678

evaluating...

Confusion matrix, without normalization

[[ 205 74 15]

[ 634 3109 78]

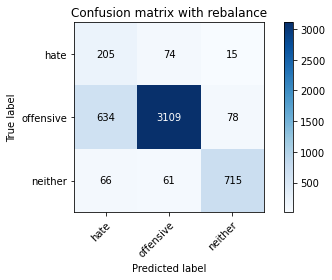
[ 66 61 715]]

Normalized confusion matrix

[[0.69727891 0.25170068 0.05102041]

[0.16592515 0.81366135 0.0204135 ]

[0.0783848 0.07244656 0.84916865]]





precision recall f1-score support

hate 0.23 0.70 0.34 294

offensive 0.96 0.81 0.88 3821

neither 0.88 0.85 0.87 842

accuracy 0.81 4957

macro avg 0.69 0.79 0.70 4957

weighted avg 0.90 0.81 0.85 4957

average eval\_loss: 0.7775744267874548, accuracy: 0.8127899939479524

training 8 epoch...

49.948015451431274 seconds used......

8 training finished! average train loss: 0.028737634077446252

total clf loss: 5.017555512516992 total adv loss: 2.068401815573452

evaluating...

Confusion matrix, without normalization

[[ 202 64 28]

[ 638 2998 185]

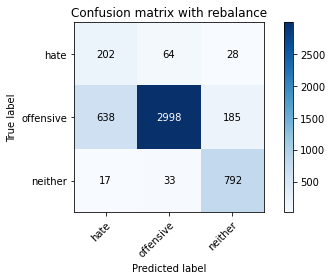
[ 17 33 792]]

Normalized confusion matrix

[[0.68707483 0.21768707 0.0952381 ]

[0.166972 0.78461136 0.04841664]

[0.02019002 0.0391924 0.94061758]]





precision recall f1-score support

hate 0.24 0.69 0.35 294

offensive 0.97 0.78 0.87 3821

neither 0.79 0.94 0.86 842

accuracy 0.81 4957

macro avg 0.66 0.80 0.69 4957

weighted avg 0.89 0.81 0.83 4957

average eval\_loss: 0.8724612275724329, accuracy: 0.8053258018963082

training 9 epoch...

49.99590063095093 seconds used......

9 training finished! average train loss: 0.011641376782024867

total clf loss: 2.072162566531915 total adv loss: 1.6297426595701836

evaluating...

Confusion matrix, without normalization

[[ 209 65 20]

[ 669 3053 99]

[ 35 59 748]]

Normalized confusion matrix

[[0.71088435 0.22108844 0.06802721]

[0.17508506 0.7990055 0.02590945]

[0.0415677 0.07007126 0.88836105]]





precision recall f1-score support

hate 0.23 0.71 0.35 294

offensive 0.96 0.80 0.87 3821

neither 0.86 0.89 0.88 842

accuracy 0.81 4957

macro avg 0.68 0.80 0.70 4957

weighted avg 0.90 0.81 0.84 4957

average eval\_loss: 0.9195528493165189, accuracy: 0.808957030461973

training 10 epoch...

49.94455337524414 seconds used......

10 training finished! average train loss: 0.017613719056812468

total clf loss: 3.048618532367982 total adv loss: 0.7334513141831849

evaluating...

Confusion matrix, without normalization

[[ 203 73 18]

[ 596 3103 122]

[ 65 28 749]]

Normalized confusion matrix

[[0.69047619 0.24829932 0.06122449]

[0.15598011 0.81209108 0.03192881]

[0.07719715 0.03325416 0.88954869]]





precision recall f1-score support

hate 0.23 0.69 0.35 294

offensive 0.97 0.81 0.88 3821

neither 0.84 0.89 0.87 842

accuracy 0.82 4957

macro avg 0.68 0.80 0.70 4957

weighted avg 0.90 0.82 0.85 4957

average eval\_loss: 0.898595847670109, accuracy: 0.8180351018761347

4 run...

training 1 epoch...

49.93894863128662 seconds used......

1 training finished! average train loss: 0.633903778087326

total clf loss: 111.06856332719326 total adv loss: 53.420354479923844

evaluating...

Confusion matrix, without normalization

[[ 210 42 42]

[ 816 2838 167]

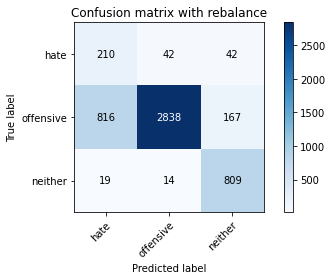
[ 19 14 809]]

Normalized confusion matrix

[[0.71428571 0.14285714 0.14285714]

[0.21355666 0.7427375 0.04370584]

[0.02256532 0.01662708 0.9608076 ]]





precision recall f1-score support

hate 0.20 0.71 0.31 294

offensive 0.98 0.74 0.85 3821

neither 0.79 0.96 0.87 842

accuracy 0.78 4957

macro avg 0.66 0.81 0.68 4957

weighted avg 0.90 0.78 0.82 4957

average eval\_loss: 0.5167979817356794, accuracy: 0.7780915876538229

training 2 epoch...

49.90053653717041 seconds used......

2 training finished! average train loss: 0.3954164337867882

total clf loss: 68.89197207987309 total adv loss: 25.515237044543028

evaluating...

Confusion matrix, without normalization

[[ 236 37 21]

[ 916 2785 120]

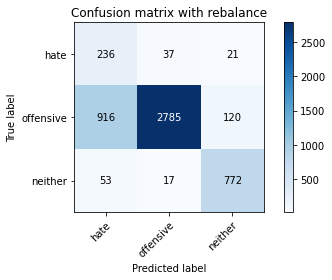
[ 53 17 772]]

Normalized confusion matrix

[[0.80272109 0.12585034 0.07142857]

[0.23972782 0.72886679 0.03140539]

[0.06294537 0.02019002 0.91686461]]





precision recall f1-score support

hate 0.20 0.80 0.31 294

offensive 0.98 0.73 0.84 3821

neither 0.85 0.92 0.88 842

accuracy 0.77 4957

macro avg 0.67 0.82 0.68 4957

weighted avg 0.91 0.77 0.81 4957

average eval\_loss: 0.525314332797162, accuracy: 0.7651805527536817

training 3 epoch...

49.882888317108154 seconds used......

3 training finished! average train loss: 0.2739032958988209

total clf loss: 47.493983410298824 total adv loss: 13.130395100219175

evaluating...

Confusion matrix, without normalization

[[ 245 30 19]

[1353 2347 121]

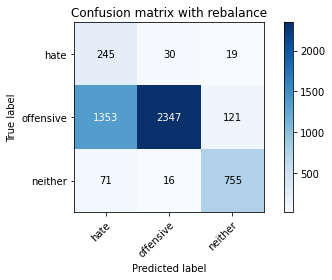
[ 71 16 755]]

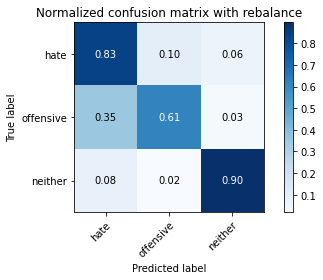
Normalized confusion matrix

[[0.83333333 0.10204082 0.06462585]

[0.35409579 0.61423711 0.0316671 ]

[0.08432304 0.01900238 0.89667458]]





precision recall f1-score support

hate 0.15 0.83 0.25 294

offensive 0.98 0.61 0.76 3821

neither 0.84 0.90 0.87 842

accuracy 0.68 4957

macro avg 0.66 0.78 0.62 4957

weighted avg 0.91 0.68 0.74 4957

average eval\_loss: 0.7789071999610432, accuracy: 0.6752067782933225

training 4 epoch...

49.908771991729736 seconds used......

4 training finished! average train loss: 0.17216362590678255

total clf loss: 29.771546041592956 total adv loss: 6.631320745393168

evaluating...

Confusion matrix, without normalization

[[ 200 80 14]

[ 457 3307 57]

[ 55 96 691]]

Normalized confusion matrix

[[0.68027211 0.27210884 0.04761905]

[0.1196022 0.86548024 0.01491756]

[0.06532067 0.11401425 0.82066508]]





precision recall f1-score support

hate 0.28 0.68 0.40 294

offensive 0.95 0.87 0.91 3821

neither 0.91 0.82 0.86 842

accuracy 0.85 4957

macro avg 0.71 0.79 0.72 4957

weighted avg 0.90 0.85 0.87 4957

average eval\_loss: 0.42438818789988514, accuracy: 0.8468831954811378

training 5 epoch...

49.88117074966431 seconds used......

5 training finished! average train loss: 0.09849751471100669

total clf loss: 17.010106842033565 total adv loss: 3.340637230227003

evaluating...

Confusion matrix, without normalization

[[ 232 48 14]

[ 887 2852 82]

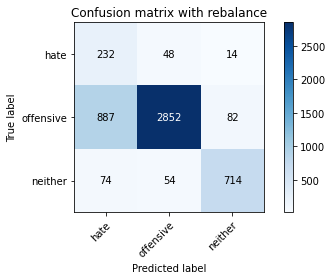
[ 74 54 714]]

Normalized confusion matrix

[[0.78911565 0.16326531 0.04761905]

[0.23213818 0.74640147 0.02146035]

[0.08788599 0.06413302 0.847981 ]]





precision recall f1-score support

hate 0.19 0.79 0.31 294

offensive 0.97 0.75 0.84 3821

neither 0.88 0.85 0.86 842

accuracy 0.77 4957

macro avg 0.68 0.79 0.67 4957

weighted avg 0.91 0.77 0.81 4957

average eval\_loss: 0.7090317296284822, accuracy: 0.7661892273552552

training 6 epoch...

49.91220498085022 seconds used......

6 training finished! average train loss: 0.0510206561236389

total clf loss: 8.837597673467826 total adv loss: 2.2613094367843587

evaluating...

Confusion matrix, without normalization

[[ 219 54 21]

[ 749 2956 116]

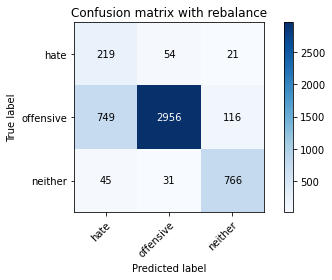
[ 45 31 766]]

Normalized confusion matrix

[[0.74489796 0.18367347 0.07142857]

[0.19602198 0.77361947 0.03035854]

[0.05344418 0.0368171 0.90973872]]





precision recall f1-score support

hate 0.22 0.74 0.34 294

offensive 0.97 0.77 0.86 3821

neither 0.85 0.91 0.88 842

accuracy 0.80 4957

macro avg 0.68 0.81 0.69 4957

weighted avg 0.91 0.80 0.83 4957

average eval\_loss: 0.9046678052719442, accuracy: 0.7950373209602583

training 7 epoch...

49.97481417655945 seconds used......

7 training finished! average train loss: 0.02257307139051702

total clf loss: 3.903583277540747 total adv loss: 0.8717613307344436

evaluating...

Confusion matrix, without normalization

[[ 193 81 20]

[ 507 3197 117]

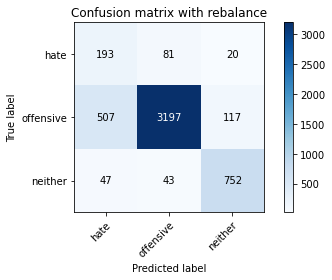
[ 47 43 752]]

Normalized confusion matrix

[[0.65646259 0.2755102 0.06802721]

[0.13268778 0.83669197 0.03062026]

[0.05581948 0.05106888 0.89311164]]





precision recall f1-score support

hate 0.26 0.66 0.37 294

offensive 0.96 0.84 0.90 3821

neither 0.85 0.89 0.87 842

accuracy 0.84 4957

macro avg 0.69 0.80 0.71 4957

weighted avg 0.90 0.84 0.86 4957

average eval\_loss: 0.7433889353227231, accuracy: 0.8355860399435142

training 8 epoch...

49.885274171829224 seconds used......

8 training finished! average train loss: 0.028251723965700738

total clf loss: 4.899199349514674 total adv loss: 1.3630911724976613

evaluating...

Confusion matrix, without normalization

[[ 220 54 20]

[ 780 2963 78]

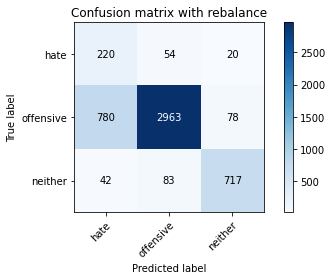
[ 42 83 717]]

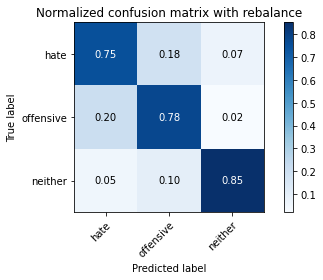
Normalized confusion matrix

[[0.74829932 0.18367347 0.06802721]

[0.20413504 0.77545145 0.0204135 ]

[0.04988124 0.09857482 0.85154394]]





precision recall f1-score support

hate 0.21 0.75 0.33 294

offensive 0.96 0.78 0.86 3821

neither 0.88 0.85 0.87 842

accuracy 0.79 4957

macro avg 0.68 0.79 0.68 4957

weighted avg 0.90 0.79 0.83 4957

average eval\_loss: 0.9442924519009408, accuracy: 0.7867661892273553

training 9 epoch...

49.925638914108276 seconds used......

9 training finished! average train loss: 0.023298414069114183

total clf loss: 4.0714341262646485 total adv loss: 1.7481064429375692

evaluating...

Confusion matrix, without normalization

[[ 245 33 16]

[1201 2521 99]

[ 89 19 734]]

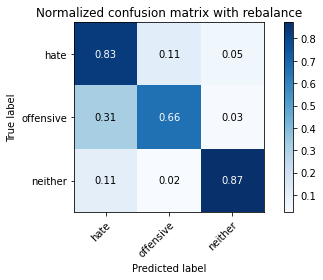
Normalized confusion matrix

[[0.83333333 0.1122449 0.05442177]

[0.31431562 0.65977493 0.02590945]

[0.10570071 0.02256532 0.87173397]]





precision recall f1-score support

hate 0.16 0.83 0.27 294

offensive 0.98 0.66 0.79 3821

neither 0.86 0.87 0.87 842

accuracy 0.71 4957

macro avg 0.67 0.79 0.64 4957

weighted avg 0.91 0.71 0.77 4957

average eval\_loss: 1.6311946379381321, accuracy: 0.7060722211014727

training 10 epoch...

49.99681878089905 seconds used......

10 training finished! average train loss: 0.01858803992158303

total clf loss: 3.2269224447663873 total adv loss: 0.9673523565579671

evaluating...

Confusion matrix, without normalization

[[ 164 109 21]

[ 329 3394 98]

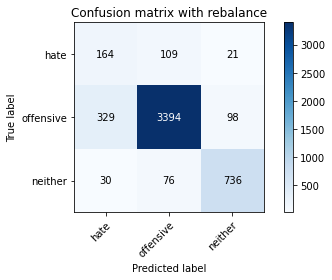
[ 30 76 736]]

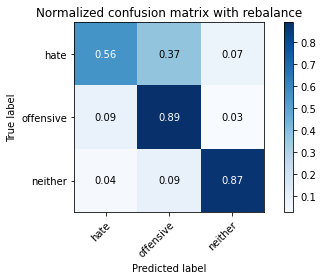
Normalized confusion matrix

[[0.55782313 0.3707483 0.07142857]

[0.08610311 0.88824915 0.02564774]

[0.03562945 0.09026128 0.87410926]]





precision recall f1-score support

hate 0.31 0.56 0.40 294

offensive 0.95 0.89 0.92 3821

neither 0.86 0.87 0.87 842

accuracy 0.87 4957

macro avg 0.71 0.77 0.73 4957

weighted avg 0.90 0.87 0.88 4957

average eval\_loss: 0.6743624808786403, accuracy: 0.8662497478313496

5 run...

training 1 epoch...

49.956043004989624 seconds used......

1 training finished! average train loss: 0.6887826567504838

total clf loss: 120.74026626348495 total adv loss: 59.168644208461046

evaluating...

Confusion matrix, without normalization

[[ 234 43 17]

[ 800 2921 100]

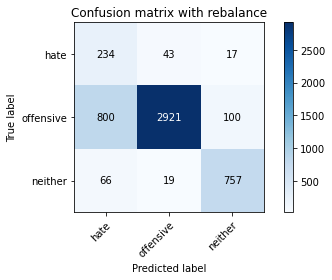
[ 66 19 757]]

Normalized confusion matrix

[[0.79591837 0.1462585 0.05782313]

[0.20936928 0.76445957 0.02617116]

[0.0783848 0.02256532 0.89904988]]





precision recall f1-score support

hate 0.21 0.80 0.34 294

offensive 0.98 0.76 0.86 3821

neither 0.87 0.90 0.88 842

accuracy 0.79 4957

macro avg 0.69 0.82 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.48141646787764564, accuracy: 0.7891870082711318

training 2 epoch...

49.936233043670654 seconds used......

2 training finished! average train loss: 0.3976346332775919

total clf loss: 69.36439402401447 total adv loss: 27.37743840366602

evaluating...

Confusion matrix, without normalization

[[ 225 50 19]

[ 742 2991 88]

[ 69 31 742]]

Normalized confusion matrix

[[0.76530612 0.17006803 0.06462585]

[0.19419 0.78277938 0.02303062]

[0.08194774 0.0368171 0.88123515]]





precision recall f1-score support

hate 0.22 0.77 0.34 294

offensive 0.97 0.78 0.87 3821

neither 0.87 0.88 0.88 842

accuracy 0.80 4957

macro avg 0.69 0.81 0.69 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.44392448062858275, accuracy: 0.7984668146056082

training 3 epoch...

49.90604877471924 seconds used......

3 training finished! average train loss: 0.269687430639016

total clf loss: 46.84998703747988 total adv loss: 14.668729827739298

evaluating...

Confusion matrix, without normalization

[[ 234 49 11]

[ 752 3008 61]

[ 94 73 675]]

Normalized confusion matrix

[[0.79591837 0.16666667 0.03741497]

[0.19680712 0.78722847 0.01596441]

[0.11163895 0.08669834 0.80166271]]





precision recall f1-score support

hate 0.22 0.80 0.34 294

offensive 0.96 0.79 0.87 3821

neither 0.90 0.80 0.85 842

accuracy 0.79 4957

macro avg 0.69 0.79 0.69 4957

weighted avg 0.91 0.79 0.83 4957

average eval\_loss: 0.49540062813508895, accuracy: 0.7901956828727053

training 4 epoch...

49.92853784561157 seconds used......

4 training finished! average train loss: 0.15205746373891482

total clf loss: 26.250508576631546 total adv loss: 4.973644635581877

evaluating...

Confusion matrix, without normalization

[[ 230 48 16]

[ 861 2800 160]

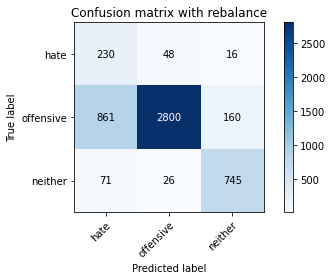
[ 71 26 745]]

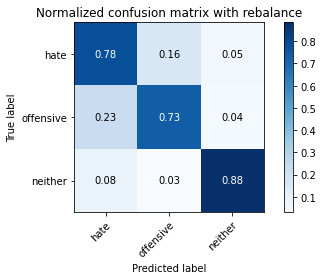
Normalized confusion matrix

[[0.78231293 0.16326531 0.05442177]

[0.22533368 0.73279246 0.04187386]

[0.08432304 0.03087886 0.8847981 ]]





precision recall f1-score support

hate 0.20 0.78 0.32 294

offensive 0.97 0.73 0.84 3821

neither 0.81 0.88 0.85 842

accuracy 0.76 4957

macro avg 0.66 0.80 0.67 4957

weighted avg 0.90 0.76 0.81 4957

average eval\_loss: 0.6383468946291795, accuracy: 0.7615493241880169

training 5 epoch...

49.88913655281067 seconds used......

5 training finished! average train loss: 0.08367033265000046

total clf loss: 14.426466019940563 total adv loss: 2.376782959414413

evaluating...

Confusion matrix, without normalization

[[ 222 61 11]

[ 702 3031 88]

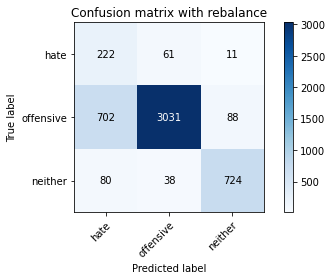
[ 80 38 724]]

Normalized confusion matrix

[[0.75510204 0.20748299 0.03741497]

[0.18372154 0.79324784 0.02303062]

[0.09501188 0.04513064 0.85985748]]





precision recall f1-score support

hate 0.22 0.76 0.34 294

offensive 0.97 0.79 0.87 3821

neither 0.88 0.86 0.87 842

accuracy 0.80 4957

macro avg 0.69 0.80 0.69 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.7048668260160353, accuracy: 0.8022997780915877

training 6 epoch...

49.95484685897827 seconds used......

6 training finished! average train loss: 0.0485521855039729

total clf loss: 8.401511286152527 total adv loss: 1.9817516269686166

evaluating...

Confusion matrix, without normalization

[[ 232 45 17]

[1062 2676 83]

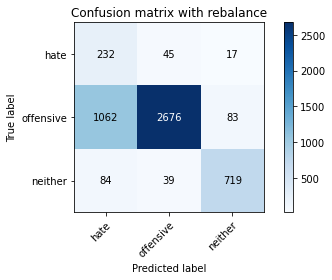
[ 84 39 719]]

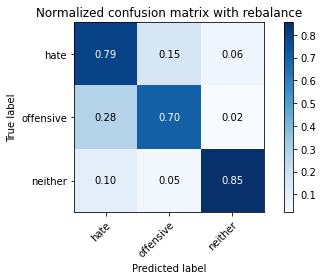
Normalized confusion matrix

[[0.78911565 0.15306122 0.05782313]

[0.27793771 0.70034023 0.02172206]

[0.09976247 0.04631829 0.85391924]]





precision recall f1-score support

hate 0.17 0.79 0.28 294

offensive 0.97 0.70 0.81 3821

neither 0.88 0.85 0.87 842

accuracy 0.73 4957

macro avg 0.67 0.78 0.65 4957

weighted avg 0.91 0.73 0.79 4957

average eval\_loss: 1.0390086540381513, accuracy: 0.7316925559814403

training 7 epoch...

49.94336414337158 seconds used......

7 training finished! average train loss: 0.03060263194578117

total clf loss: 5.263439217291307 total adv loss: 0.6077829851783463

evaluating...

Confusion matrix, without normalization

[[ 233 51 10]

[ 812 2946 63]

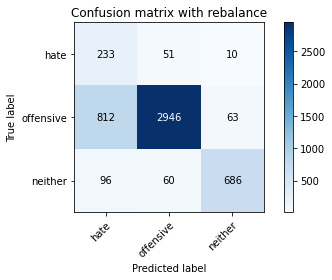
[ 96 60 686]]

Normalized confusion matrix

[[0.79251701 0.17346939 0.03401361]

[0.21250981 0.77100236 0.01648783]

[0.11401425 0.07125891 0.81472684]]





precision recall f1-score support

hate 0.20 0.79 0.32 294

offensive 0.96 0.77 0.86 3821

neither 0.90 0.81 0.86 842

accuracy 0.78 4957

macro avg 0.69 0.79 0.68 4957

weighted avg 0.91 0.78 0.83 4957

average eval\_loss: 1.00897965445033, accuracy: 0.7797054670163406

training 8 epoch...

49.98455762863159 seconds used......

8 training finished! average train loss: 0.018676422780947433

total clf loss: 3.28337307649781 total adv loss: 1.7940958680264885

evaluating...

Confusion matrix, without normalization

[[ 233 44 17]

[ 917 2786 118]

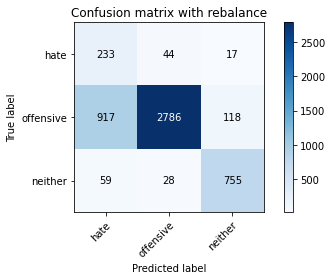
[ 59 28 755]]

Normalized confusion matrix

[[0.79251701 0.14965986 0.05782313]

[0.23998953 0.7291285 0.03088197]

[0.07007126 0.03325416 0.89667458]]





precision recall f1-score support

hate 0.19 0.79 0.31 294

offensive 0.97 0.73 0.83 3821

neither 0.85 0.90 0.87 842

accuracy 0.76 4957

macro avg 0.67 0.81 0.67 4957

weighted avg 0.91 0.76 0.81 4957

average eval\_loss: 1.1649062229816112, accuracy: 0.7613475892677022

training 9 epoch...

49.91901445388794 seconds used......

9 training finished! average train loss: 0.029055858395734328

total clf loss: 5.039437877945602 total adv loss: 1.4177218050390366

evaluating...

Confusion matrix, without normalization

[[ 210 64 20]

[ 592 3121 108]

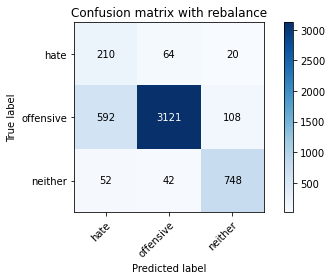
[ 52 42 748]]

Normalized confusion matrix

[[0.71428571 0.21768707 0.06802721]

[0.15493326 0.81680188 0.02826485]

[0.06175772 0.04988124 0.88836105]]





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 0.91 0.82 0.85 4957

average eval\_loss: 0.845164148628922, accuracy: 0.8228767399636877

training 10 epoch...

49.88416409492493 seconds used......

10 training finished! average train loss: 0.01344728149383122

total clf loss: 2.306745262379991 total adv loss: 0.145202190404234

evaluating...

Confusion matrix, without normalization

[[ 209 61 24]

[ 545 3133 143]

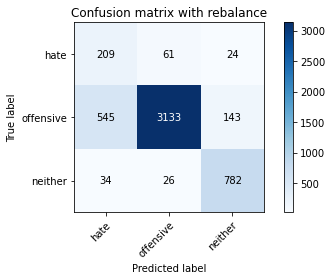
[ 34 26 782]]

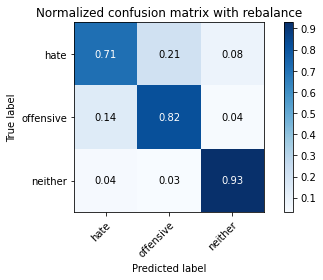
Normalized confusion matrix

[[0.71088435 0.20748299 0.08163265]

[0.14263282 0.81994242 0.03742476]

[0.04038005 0.03087886 0.92874109]]





precision recall f1-score support

hate 0.27 0.71 0.39 294

offensive 0.97 0.82 0.89 3821

neither 0.82 0.93 0.87 842

accuracy 0.83 4957

macro avg 0.69 0.82 0.72 4957

weighted avg 0.91 0.83 0.86 4957

average eval\_loss: 0.912538751010523, accuracy: 0.8319548113778495