training 1 epoch...

44.976950883865356 seconds used......

1 training finished! average train loss: 1.2608554500933977

total clf loss: 115.78279502689838 total adv loss: 71.30249160155654

evaluating...

Confusion matrix, without normalization

[[ 244 35 15]

[1060 2683 78]

[ 122 22 698]]

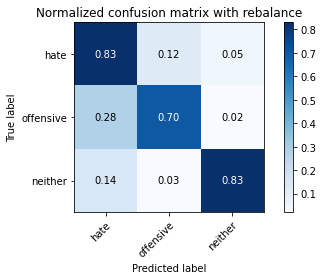
Normalized confusion matrix

[[0.82993197 0.11904762 0.05102041]

[0.27741429 0.70217221 0.0204135 ]

[0.14489311 0.02612827 0.82897862]]





precision recall f1-score support

hate 0.17 0.83 0.28 294

offensive 0.98 0.70 0.82 3821

neither 0.88 0.83 0.85 842

accuracy 0.73 4957

macro avg 0.68 0.79 0.65 4957

weighted avg 0.91 0.73 0.79 4957

average eval\_loss: 0.6702386035674042, accuracy: 0.731289086140811

evaluating adversary...

training 2 epoch...

45.008429765701294 seconds used......

2 training finished! average train loss: 0.8233772380310193

total clf loss: 71.79413664340973 total adv loss: 49.28812316432595

evaluating...

Confusion matrix, without normalization

[[ 227 44 23]

[ 729 2975 117]

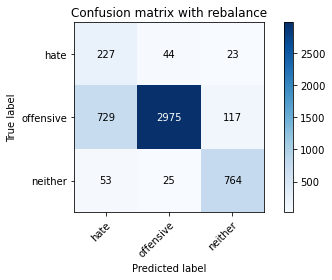
[ 53 25 764]]

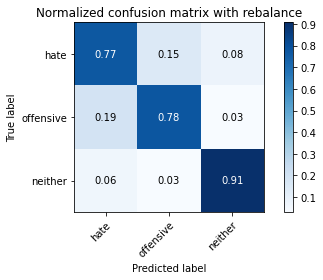
Normalized confusion matrix

[[0.77210884 0.14965986 0.07823129]

[0.19078775 0.77859199 0.03062026]

[0.06294537 0.02969121 0.90736342]]





precision recall f1-score support

hate 0.22 0.77 0.35 294

offensive 0.98 0.78 0.87 3821

neither 0.85 0.91 0.88 842

accuracy 0.80 4957

macro avg 0.68 0.82 0.70 4957

weighted avg 0.91 0.80 0.84 4957

average eval\_loss: 0.4618401541825264, accuracy: 0.8000806939681259

evaluating adversary...

training 3 epoch...

44.944867849349976 seconds used......

3 training finished! average train loss: 0.6258885562419891

total clf loss: 52.89105557650328 total adv loss: 38.66849138960242

evaluating...

Confusion matrix, without normalization

[[ 229 44 21]

[ 787 2810 224]

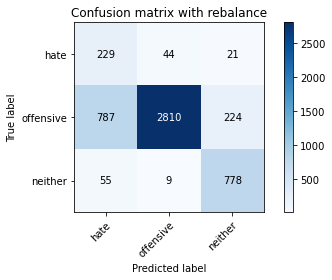
[ 55 9 778]]

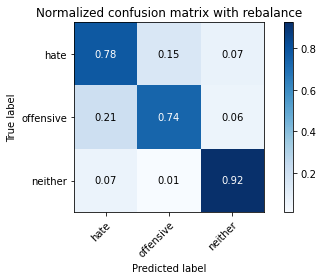
Normalized confusion matrix

[[0.77891156 0.14965986 0.07142857]

[0.20596702 0.73540958 0.0586234 ]

[0.06532067 0.01068884 0.9239905 ]]





precision recall f1-score support

hate 0.21 0.78 0.34 294

offensive 0.98 0.74 0.84 3821

neither 0.76 0.92 0.83 842

accuracy 0.77 4957

macro avg 0.65 0.81 0.67 4957

weighted avg 0.90 0.77 0.81 4957

average eval\_loss: 0.5774612760832233, accuracy: 0.7700221908412346

evaluating adversary...

training 4 epoch...

44.85209250450134 seconds used......

4 training finished! average train loss: 0.36857587059861735

total clf loss: 30.386273816227913 total adv loss: 23.31442849524319

evaluating...

Confusion matrix, without normalization

[[ 213 63 18]

[ 657 3024 140]

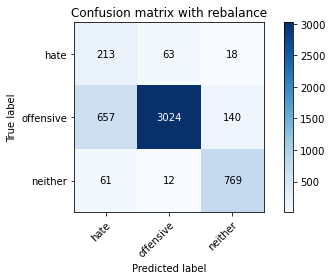
[ 61 12 769]]

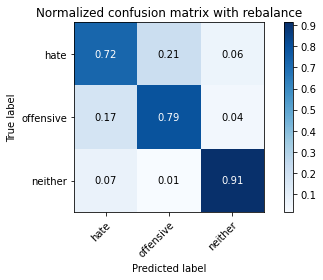
Normalized confusion matrix

[[0.7244898 0.21428571 0.06122449]

[0.17194452 0.79141586 0.03663962]

[0.07244656 0.01425178 0.91330166]]





precision recall f1-score support

hate 0.23 0.72 0.35 294

offensive 0.98 0.79 0.87 3821

neither 0.83 0.91 0.87 842

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: 0.597800794057548, accuracy: 0.8081500907807141

evaluating adversary...

training 5 epoch...

44.90160536766052 seconds used......

5 training finished! average train loss: 0.1786346557381296

total clf loss: 14.383288765791804 total adv loss: 11.545169740740675

evaluating...

Confusion matrix, without normalization

[[ 221 56 17]

[ 848 2874 99]

[ 68 44 730]]

Normalized confusion matrix

[[0.75170068 0.19047619 0.05782313]

[0.22193143 0.75215912 0.02590945]

[0.0807601 0.05225653 0.86698337]]





precision recall f1-score support

hate 0.19 0.75 0.31 294

offensive 0.97 0.75 0.85 3821

neither 0.86 0.87 0.86 842

accuracy 0.77 4957

macro avg 0.67 0.79 0.67 4957

weighted avg 0.90 0.77 0.82 4957

average eval\_loss: 0.7936273375450964, accuracy: 0.7716360702037522

evaluating adversary...

training 6 epoch...

45.0298969745636 seconds used......

6 training finished! average train loss: 0.11693314407075271

total clf loss: 9.432211751118302 total adv loss: 7.545254272656166

evaluating...

Confusion matrix, without normalization

[[ 207 69 18]

[ 615 3101 105]

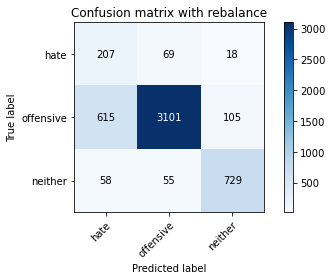
[ 58 55 729]]

Normalized confusion matrix

[[0.70408163 0.23469388 0.06122449]

[0.16095263 0.81156765 0.02747972]

[0.06888361 0.06532067 0.86579572]]





precision recall f1-score support

hate 0.24 0.70 0.35 294

offensive 0.96 0.81 0.88 3821

neither 0.86 0.87 0.86 842

accuracy 0.81 4957

macro avg 0.68 0.79 0.70 4957

weighted avg 0.90 0.81 0.85 4957

average eval\_loss: 0.738294776780681, accuracy: 0.81440387331047

evaluating adversary...

training 7 epoch...

44.88066744804382 seconds used......

7 training finished! average train loss: 0.07628448153144751

total clf loss: 6.059935746947303 total adv loss: 4.989079066035629

evaluating...

Confusion matrix, without normalization

[[ 215 54 25]

[ 690 2981 150]

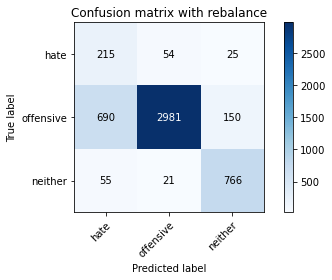
[ 55 21 766]]

Normalized confusion matrix

[[0.73129252 0.18367347 0.08503401]

[0.180581 0.78016226 0.03925674]

[0.06532067 0.02494062 0.90973872]]





precision recall f1-score support

hate 0.22 0.73 0.34 294

offensive 0.98 0.78 0.87 3821

neither 0.81 0.91 0.86 842

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: 0.9165410848236794, accuracy: 0.7992737542868671

evaluating adversary...

training 8 epoch...

45.0446617603302 seconds used......

8 training finished! average train loss: 0.045157208682680555

total clf loss: 3.581349754764233 total adv loss: 2.9575235737429466

evaluating...

Confusion matrix, without normalization

[[ 197 76 21]

[ 498 3205 118]

[ 47 48 747]]

Normalized confusion matrix

[[0.67006803 0.2585034 0.07142857]

[0.13033237 0.83878566 0.03088197]

[0.05581948 0.05700713 0.8871734 ]]





precision recall f1-score support

hate 0.27 0.67 0.38 294

offensive 0.96 0.84 0.90 3821

neither 0.84 0.89 0.86 842

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: 0.7924619322871994, accuracy: 0.8369981843857172

evaluating adversary...

training 9 epoch...

44.924968957901 seconds used......

9 training finished! average train loss: 0.07935572406910196

total clf loss: 6.4051882199710235 total adv loss: 5.117600508136093

evaluating...

Confusion matrix, without normalization

[[ 224 49 21]

[ 780 2886 155]

[ 53 26 763]]

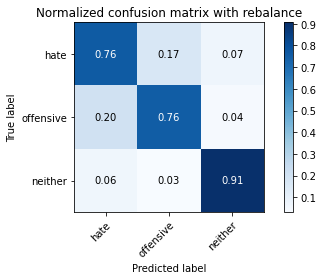
Normalized confusion matrix

[[0.76190476 0.16666667 0.07142857]

[0.20413504 0.75529966 0.0405653 ]

[0.06294537 0.03087886 0.90617577]]





precision recall f1-score support

hate 0.21 0.76 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 0.81 0.68 4957

weighted avg 0.90 0.78 0.82 4957

average eval\_loss: 1.0093240651542381, accuracy: 0.7813193463788581

evaluating adversary...

training 10 epoch...

44.95928478240967 seconds used......

10 training finished! average train loss: 0.029696309472684753

total clf loss: 2.263159921625629 total adv loss: 2.010649298959379

evaluating...

Confusion matrix, without normalization

[[ 240 39 15]

[ 871 2825 125]

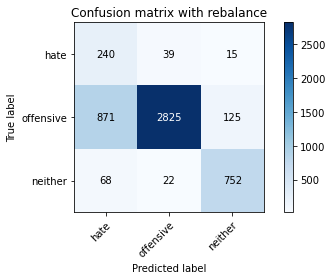
[ 68 22 752]]

Normalized confusion matrix

[[0.81632653 0.13265306 0.05102041]

[0.2279508 0.73933525 0.03271395]

[0.0807601 0.02612827 0.89311164]]





precision recall f1-score support

hate 0.20 0.82 0.33 294

offensive 0.98 0.74 0.84 3821

neither 0.84 0.89 0.87 842

accuracy 0.77 4957

macro avg 0.68 0.82 0.68 4957

weighted avg 0.91 0.77 0.82 4957

average eval\_loss: 1.297552330858795, accuracy: 0.7700221908412346

evaluating adversary...