1 run......

Downloading: 100%

464/464 [00:00<00:00, 13.2kB/s]

Downloading: 100%

270M/270M [00:09<00:00, 29.2MB/s]

training 1 epoch...

18.689752101898193 seconds used......

1 training finished! average train loss: 1.8235563975114089

total clf loss: 51.747087240219116 total adv loss: 33.39203943312168

evaluating...

Confusion matrix, without normalization

[[ 0 0 41]

[ 0 0 95]

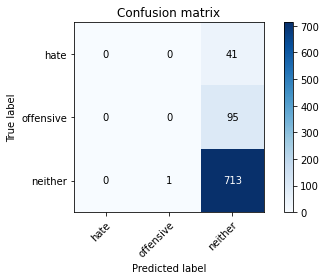
[ 0 1 713]]

Normalized confusion matrix

[[0. 0. 1. ]

[0. 0. 1. ]

[0. 0.00140056 0.99859944]]





/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.00 0.00 0.00 95

neither 0.84 1.00 0.91 714

accuracy 0.84 850

macro avg 0.28 0.33 0.30 850

weighted avg 0.71 0.84 0.77 850

average eval\_loss: 0.5576935247626416, accuracy: 0.8388235294117647

training 2 epoch...

17.703295946121216 seconds used......

2 training finished! average train loss: 1.4912057087971613

total clf loss: 43.336094319820404 total adv loss: 26.796138375997543

evaluating...

Confusion matrix, without normalization

[[ 0 9 32]

[ 0 32 63]

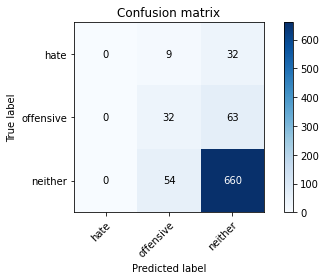
[ 0 54 660]]

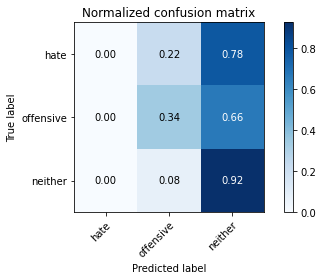
Normalized confusion matrix

[[0. 0.2195122 0.7804878 ]

[0. 0.33684211 0.66315789]

[0. 0.07563025 0.92436975]]





/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.34 0.34 0.34 95

neither 0.87 0.92 0.90 714

accuracy 0.81 850

macro avg 0.40 0.42 0.41 850

weighted avg 0.77 0.81 0.79 850

average eval\_loss: 0.5439901651251454, accuracy: 0.8141176470588235

training 3 epoch...

17.60386562347412 seconds used......

3 training finished! average train loss: 1.12699174789282

total clf loss: 33.728242844343185 total adv loss: 19.76311033219099

evaluating...

Confusion matrix, without normalization

[[ 0 11 30]

[ 1 39 55]

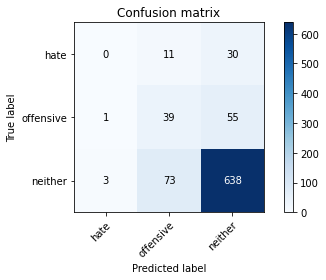
[ 3 73 638]]

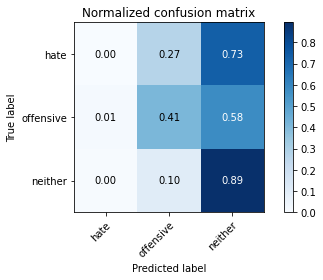
Normalized confusion matrix

[[0. 0.26829268 0.73170732]

[0.01052632 0.41052632 0.57894737]

[0.00420168 0.1022409 0.89355742]]





precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.32 0.41 0.36 95

neither 0.88 0.89 0.89 714

accuracy 0.80 850

macro avg 0.40 0.43 0.42 850

weighted avg 0.78 0.80 0.79 850

average eval\_loss: 0.5487399136777534, accuracy: 0.7964705882352942

training 4 epoch...

17.763978719711304 seconds used......

4 training finished! average train loss: 0.5727304270634285

total clf loss: 21.94904015213251 total adv loss: 7.639218684285879

evaluating...

Confusion matrix, without normalization

[[ 1 13 27]

[ 2 51 42]

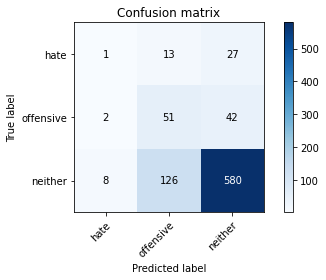
[ 8 126 580]]

Normalized confusion matrix

[[0.02439024 0.31707317 0.65853659]

[0.02105263 0.53684211 0.44210526]

[0.01120448 0.17647059 0.81232493]]





precision recall f1-score support

hate 0.09 0.02 0.04 41

offensive 0.27 0.54 0.36 95

neither 0.89 0.81 0.85 714

accuracy 0.74 850

macro avg 0.42 0.46 0.42 850

weighted avg 0.79 0.74 0.76 850

average eval\_loss: 0.6748278861388911, accuracy: 0.7435294117647059

training 5 epoch...

17.67205786705017 seconds used......

5 training finished! average train loss: 0.32637182754965927

total clf loss: 14.703617807477713 total adv loss: 3.255275465664454

evaluating...

Confusion matrix, without normalization

[[ 5 8 28]

[ 5 39 51]

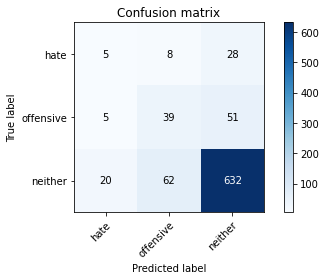
[ 20 62 632]]

Normalized confusion matrix

[[0.12195122 0.19512195 0.68292683]

[0.05263158 0.41052632 0.53684211]

[0.0280112 0.08683473 0.88515406]]





precision recall f1-score support

hate 0.17 0.12 0.14 41

offensive 0.36 0.41 0.38 95

neither 0.89 0.89 0.89 714

accuracy 0.80 850

macro avg 0.47 0.47 0.47 850

weighted avg 0.79 0.80 0.79 850

average eval\_loss: 0.6257717797944192, accuracy: 0.7952941176470588

training 6 epoch...

17.637094497680664 seconds used......

6 training finished! average train loss: 0.22529460684611247

total clf loss: 11.112761616706848 total adv loss: 1.7656938774962327

evaluating...

Confusion matrix, without normalization

[[ 11 10 20]

[ 6 65 24]

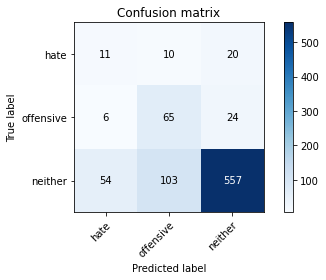
[ 54 103 557]]

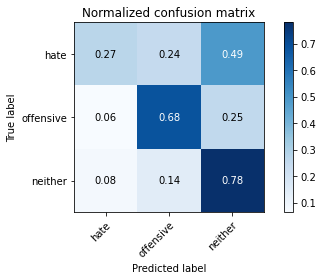
Normalized confusion matrix

[[0.26829268 0.24390244 0.48780488]

[0.06315789 0.68421053 0.25263158]

[0.07563025 0.1442577 0.78011204]]





precision recall f1-score support

hate 0.15 0.27 0.20 41

offensive 0.37 0.68 0.48 95

neither 0.93 0.78 0.85 714

accuracy 0.74 850

macro avg 0.48 0.58 0.51 850

weighted avg 0.83 0.74 0.77 850

average eval\_loss: 0.7579493986139464, accuracy: 0.7447058823529412

training 7 epoch...

17.54721689224243 seconds used......

7 training finished! average train loss: 0.1714396972782337

total clf loss: 8.504315007477999 total adv loss: 1.319632628794352

evaluating...

Confusion matrix, without normalization

[[ 10 7 24]

[ 9 48 38]

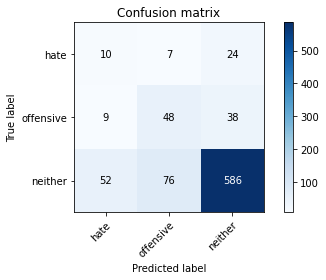
[ 52 76 586]]

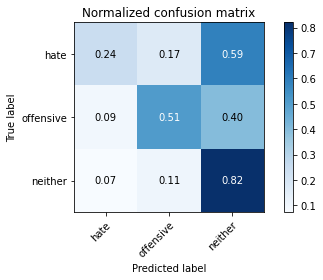
Normalized confusion matrix

[[0.24390244 0.17073171 0.58536585]

[0.09473684 0.50526316 0.4 ]

[0.07282913 0.10644258 0.82072829]]





precision recall f1-score support

hate 0.14 0.24 0.18 41

offensive 0.37 0.51 0.42 95

neither 0.90 0.82 0.86 714

accuracy 0.76 850

macro avg 0.47 0.52 0.49 850

weighted avg 0.81 0.76 0.78 850

average eval\_loss: 0.7841775860355864, accuracy: 0.7576470588235295

training 8 epoch...

17.75130271911621 seconds used......

8 training finished! average train loss: 0.09907122010794969

total clf loss: 6.0702572194859385 total adv loss: 0.18468603947530937

evaluating...

Confusion matrix, without normalization

[[ 16 7 18]

[ 11 45 39]

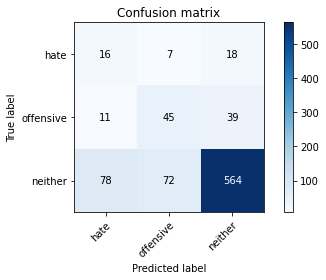
[ 78 72 564]]

Normalized confusion matrix

[[0.3902439 0.17073171 0.43902439]

[0.11578947 0.47368421 0.41052632]

[0.1092437 0.10084034 0.78991597]]





precision recall f1-score support

hate 0.15 0.39 0.22 41

offensive 0.36 0.47 0.41 95

neither 0.91 0.79 0.84 714

accuracy 0.74 850

macro avg 0.47 0.55 0.49 850

weighted avg 0.81 0.74 0.77 850

average eval\_loss: 0.9084629384302643, accuracy: 0.7352941176470589

training 9 epoch...

17.60210943222046 seconds used......

9 training finished! average train loss: 0.09481530517626267

total clf loss: 4.676997667178512 total adv loss: 0.7429986032520901

evaluating...

Confusion matrix, without normalization

[[ 18 4 19]

[ 15 24 56]

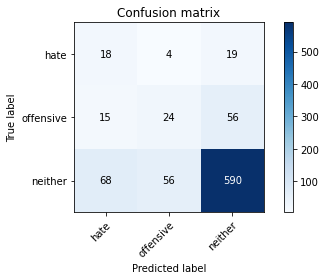
[ 68 56 590]]

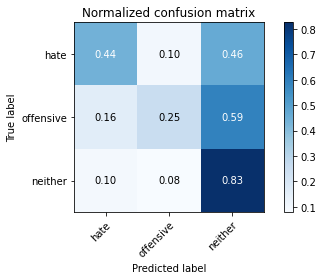
Normalized confusion matrix

[[0.43902439 0.09756098 0.46341463]

[0.15789474 0.25263158 0.58947368]

[0.0952381 0.07843137 0.82633053]]





precision recall f1-score support

hate 0.18 0.44 0.25 41

offensive 0.29 0.25 0.27 95

neither 0.89 0.83 0.86 714

accuracy 0.74 850

macro avg 0.45 0.51 0.46 850

weighted avg 0.79 0.74 0.76 850

average eval\_loss: 0.9972480512711458, accuracy: 0.7435294117647059

training 10 epoch...

17.593172073364258 seconds used......

10 training finished! average train loss: 0.12836942729325249

total clf loss: 4.6157490205951035 total adv loss: 1.8641318666905136

evaluating...

Confusion matrix, without normalization

[[ 12 7 22]

[ 12 52 31]

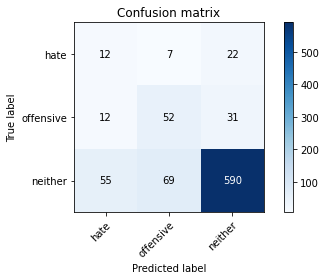
[ 55 69 590]]

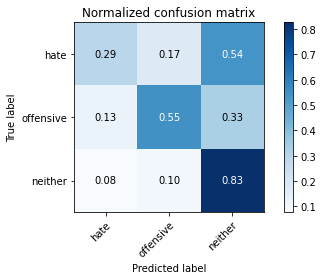
Normalized confusion matrix

[[0.29268293 0.17073171 0.53658537]

[0.12631579 0.54736842 0.32631579]

[0.07703081 0.09663866 0.82633053]]





precision recall f1-score support

hate 0.15 0.29 0.20 41

offensive 0.41 0.55 0.47 95

neither 0.92 0.83 0.87 714

accuracy 0.77 850

macro avg 0.49 0.56 0.51 850

weighted avg 0.82 0.77 0.79 850

average eval\_loss: 0.8310455991940703, accuracy: 0.7694117647058824

2 run......

training 1 epoch...

17.76619029045105 seconds used......

1 training finished! average train loss: 1.822762275659121

total clf loss: 51.9382461309433 total adv loss: 33.270650908350945

evaluating...

Confusion matrix, without normalization

[[ 0 0 41]

[ 0 0 95]

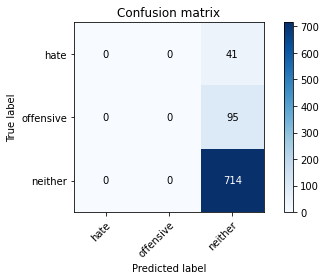
[ 0 0 714]]

Normalized confusion matrix

[[0. 0. 1.]

[0. 0. 1.]

[0. 0. 1.]]





/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.00 0.00 0.00 95

neither 0.84 1.00 0.91 714

accuracy 0.84 850

macro avg 0.28 0.33 0.30 850

weighted avg 0.71 0.84 0.77 850

average eval\_loss: 0.5972756383030914, accuracy: 0.84

training 2 epoch...

17.65133500099182 seconds used......

2 training finished! average train loss: 1.5250436975405766

total clf loss: 44.71621876955032 total adv loss: 27.205810695886612

evaluating...

Confusion matrix, without normalization

[[ 0 1 40]

[ 0 14 81]

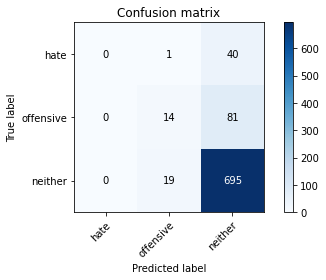
[ 0 19 695]]

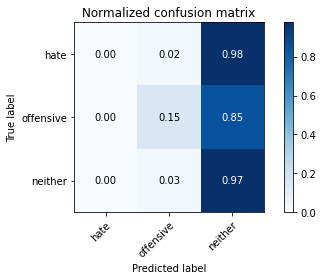
Normalized confusion matrix

[[0. 0.02439024 0.97560976]

[0. 0.14736842 0.85263158]

[0. 0.02661064 0.97338936]]





/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.41 0.15 0.22 95

neither 0.85 0.97 0.91 714

accuracy 0.83 850

macro avg 0.42 0.37 0.38 850

weighted avg 0.76 0.83 0.79 850

average eval\_loss: 0.5211573229279629, accuracy: 0.8341176470588235

training 3 epoch...

17.586330890655518 seconds used......

3 training finished! average train loss: 1.1065035411944757

total clf loss: 34.22309437394142 total adv loss: 18.84981796145439

evaluating...

Confusion matrix, without normalization

[[ 0 1 40]

[ 1 8 86]

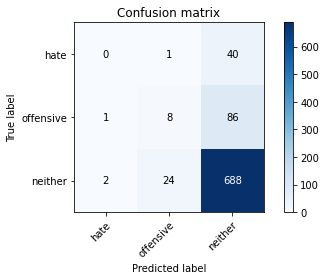
[ 2 24 688]]

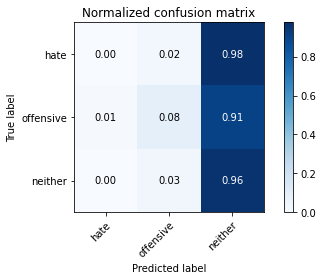
Normalized confusion matrix

[[0. 0.02439024 0.97560976]

[0.01052632 0.08421053 0.90526316]

[0.00280112 0.03361345 0.96358543]]





precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.24 0.08 0.12 95

neither 0.85 0.96 0.90 714

accuracy 0.82 850

macro avg 0.36 0.35 0.34 850

weighted avg 0.74 0.82 0.77 850

average eval\_loss: 0.532284383597069, accuracy: 0.8188235294117647

training 4 epoch...

17.670642852783203 seconds used......

4 training finished! average train loss: 0.6283109009265899

total clf loss: 22.933981880545616 total adv loss: 8.953113358002156

evaluating...

Confusion matrix, without normalization

[[ 2 16 23]

[ 5 55 35]

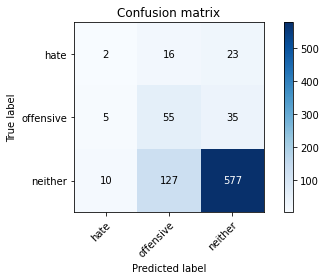
[ 10 127 577]]

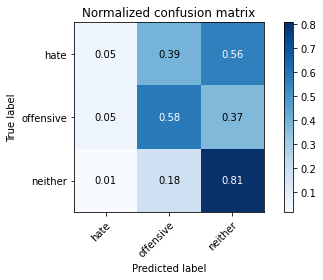
Normalized confusion matrix

[[0.04878049 0.3902439 0.56097561]

[0.05263158 0.57894737 0.36842105]

[0.0140056 0.17787115 0.80812325]]





precision recall f1-score support

hate 0.12 0.05 0.07 41

offensive 0.28 0.58 0.38 95

neither 0.91 0.81 0.86 714

accuracy 0.75 850

macro avg 0.43 0.48 0.43 850

weighted avg 0.80 0.75 0.76 850

average eval\_loss: 0.6794060768862796, accuracy: 0.7458823529411764

training 5 epoch...

17.56625485420227 seconds used......

5 training finished! average train loss: 0.3864823134472737

total clf loss: 16.620242308825254 total adv loss: 4.250554035883397

evaluating...

Confusion matrix, without normalization

[[ 2 6 33]

[ 2 28 65]

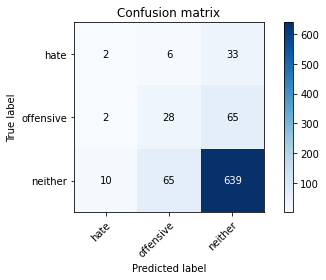
[ 10 65 639]]

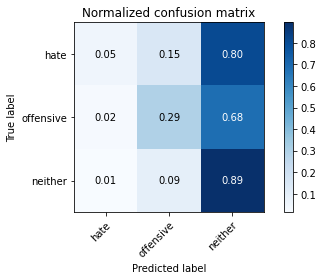
Normalized confusion matrix

[[0.04878049 0.14634146 0.80487805]

[0.02105263 0.29473684 0.68421053]

[0.0140056 0.09103641 0.89495798]]





precision recall f1-score support

hate 0.14 0.05 0.07 41

offensive 0.28 0.29 0.29 95

neither 0.87 0.89 0.88 714

accuracy 0.79 850

macro avg 0.43 0.41 0.41 850

weighted avg 0.77 0.79 0.78 850

average eval\_loss: 0.6176115571586198, accuracy: 0.7870588235294118

training 6 epoch...

17.603355646133423 seconds used......

6 training finished! average train loss: 0.19354433388664172

total clf loss: 10.738260388374329 total adv loss: 0.9210606583292247

evaluating...

Confusion matrix, without normalization

[[ 10 4 27]

[ 10 23 62]

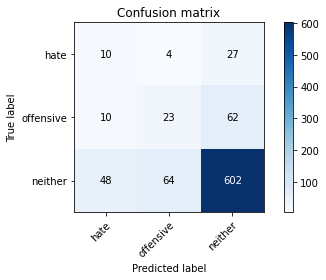
[ 48 64 602]]

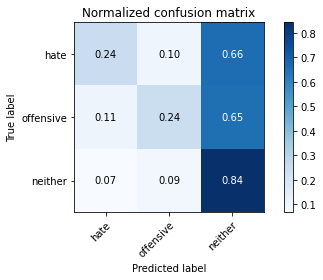
Normalized confusion matrix

[[0.24390244 0.09756098 0.65853659]

[0.10526316 0.24210526 0.65263158]

[0.06722689 0.08963585 0.84313725]]





precision recall f1-score support

hate 0.15 0.24 0.18 41

offensive 0.25 0.24 0.25 95

neither 0.87 0.84 0.86 714

accuracy 0.75 850

macro avg 0.42 0.44 0.43 850

weighted avg 0.77 0.75 0.76 850

average eval\_loss: 0.7662286059064574, accuracy: 0.7470588235294118

training 7 epoch...

17.675552129745483 seconds used......

7 training finished! average train loss: 0.154756983426901

total clf loss: 8.306928064674139 total adv loss: 0.8761379199204384

evaluating...

Confusion matrix, without normalization

[[ 5 12 24]

[ 4 37 54]

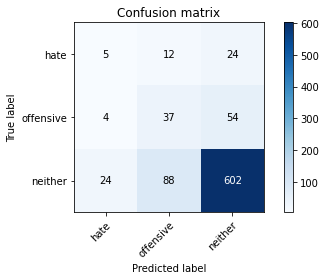
[ 24 88 602]]

Normalized confusion matrix

[[0.12195122 0.29268293 0.58536585]

[0.04210526 0.38947368 0.56842105]

[0.03361345 0.1232493 0.84313725]]





precision recall f1-score support

hate 0.15 0.12 0.14 41

offensive 0.27 0.39 0.32 95

neither 0.89 0.84 0.86 714

accuracy 0.76 850

macro avg 0.44 0.45 0.44 850

weighted avg 0.78 0.76 0.77 850

average eval\_loss: 0.8392807989446229, accuracy: 0.7576470588235295

training 8 epoch...

17.588199377059937 seconds used......

8 training finished! average train loss: 0.08972086725899807

total clf loss: 5.119556597433984 total adv loss: 0.35614988922134216

evaluating...

Confusion matrix, without normalization

[[ 11 4 26]

[ 12 26 57]

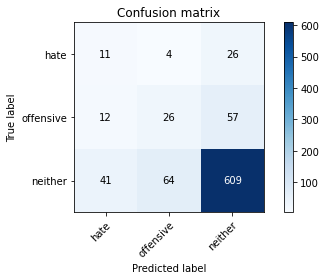
[ 41 64 609]]

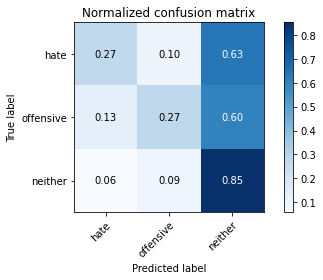
Normalized confusion matrix

[[0.26829268 0.09756098 0.63414634]

[0.12631579 0.27368421 0.6 ]

[0.05742297 0.08963585 0.85294118]]





precision recall f1-score support

hate 0.17 0.27 0.21 41

offensive 0.28 0.27 0.28 95

neither 0.88 0.85 0.87 714

accuracy 0.76 850

macro avg 0.44 0.46 0.45 850

weighted avg 0.78 0.76 0.77 850

average eval\_loss: 0.8629586798420479, accuracy: 0.76

training 9 epoch...

17.577268838882446 seconds used......

9 training finished! average train loss: 0.04900836530499733

total clf loss: 3.112008727155626 total adv loss: 0.036767508811749394

evaluating...

Confusion matrix, without normalization

[[ 14 4 23]

[ 13 26 56]

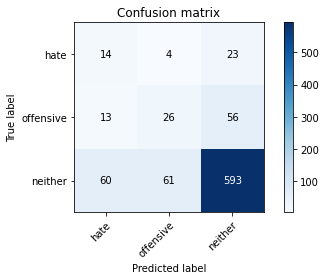
[ 60 61 593]]

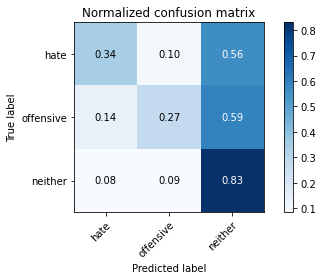
Normalized confusion matrix

[[0.34146341 0.09756098 0.56097561]

[0.13684211 0.27368421 0.58947368]

[0.08403361 0.08543417 0.83053221]]





precision recall f1-score support

hate 0.16 0.34 0.22 41

offensive 0.29 0.27 0.28 95

neither 0.88 0.83 0.86 714

accuracy 0.74 850

macro avg 0.44 0.48 0.45 850

weighted avg 0.78 0.74 0.76 850

average eval\_loss: 0.9881543180861965, accuracy: 0.7447058823529412

training 10 epoch...

17.695573329925537 seconds used......

10 training finished! average train loss: 0.08754577006285007

total clf loss: 2.8723222222179174 total adv loss: 1.4090764118973311

evaluating...

Confusion matrix, without normalization

[[ 5 7 29]

[ 10 31 54]

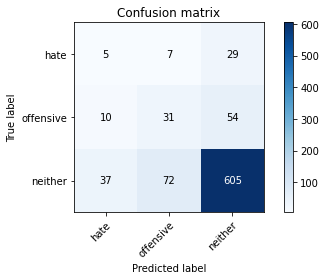
[ 37 72 605]]

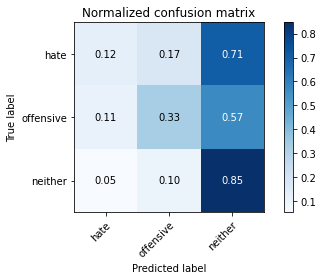
Normalized confusion matrix

[[0.12195122 0.17073171 0.70731707]

[0.10526316 0.32631579 0.56842105]

[0.05182073 0.10084034 0.84733894]]





precision recall f1-score support

hate 0.10 0.12 0.11 41

offensive 0.28 0.33 0.30 95

neither 0.88 0.85 0.86 714

accuracy 0.75 850

macro avg 0.42 0.43 0.42 850

weighted avg 0.77 0.75 0.76 850

average eval\_loss: 0.8661089913735469, accuracy: 0.7541176470588236

3 run......

training 1 epoch...

17.707926988601685 seconds used......

1 training finished! average train loss: 1.8333767542472252

total clf loss: 53.1035051047802 total adv loss: 33.03299215435982

evaluating...

Confusion matrix, without normalization

[[ 0 0 41]

[ 0 0 95]

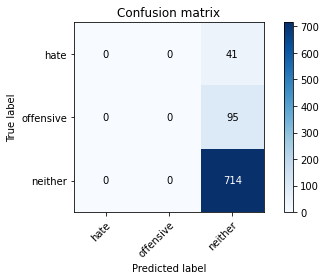
[ 0 0 714]]

Normalized confusion matrix

[[0. 0. 1.]

[0. 0. 1.]

[0. 0. 1.]]





/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.00 0.00 0.00 95

neither 0.84 1.00 0.91 714

accuracy 0.84 850

macro avg 0.28 0.33 0.30 850

weighted avg 0.71 0.84 0.77 850

average eval\_loss: 0.5794488714877949, accuracy: 0.84

training 2 epoch...

17.57645893096924 seconds used......

2 training finished! average train loss: 1.5597751030555138

total clf loss: 45.58733552694321 total adv loss: 27.899022981524467

evaluating...

Confusion matrix, without normalization

[[ 0 0 41]

[ 0 1 94]

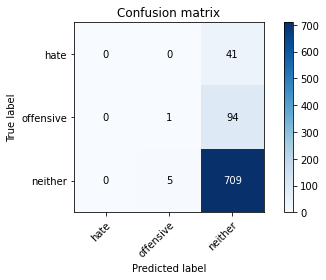
[ 0 5 709]]

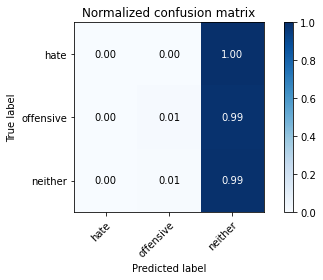
Normalized confusion matrix

[[0. 0. 1. ]

[0. 0.01052632 0.98947368]

[0. 0.0070028 0.9929972 ]]





/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.17 0.01 0.02 95

neither 0.84 0.99 0.91 714

accuracy 0.84 850

macro avg 0.34 0.33 0.31 850

weighted avg 0.72 0.84 0.77 850

average eval\_loss: 0.5172931280766808, accuracy: 0.8352941176470589

training 3 epoch...

17.66908836364746 seconds used......

3 training finished! average train loss: 1.1630231155798985

total clf loss: 35.75831858813763 total adv loss: 19.91909220814705

evaluating...

Confusion matrix, without normalization

[[ 0 7 34]

[ 0 18 77]

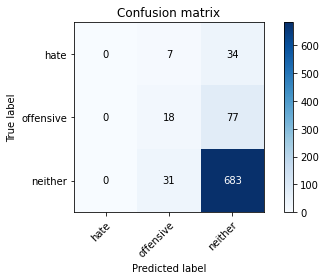
[ 0 31 683]]

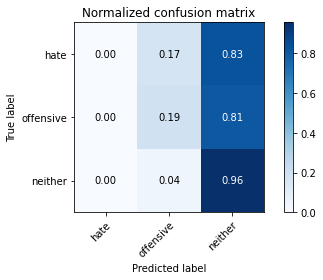
Normalized confusion matrix

[[0. 0.17073171 0.82926829]

[0. 0.18947368 0.81052632]

[0. 0.04341737 0.95658263]]





/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.32 0.19 0.24 95

neither 0.86 0.96 0.91 714

accuracy 0.82 850

macro avg 0.39 0.38 0.38 850

weighted avg 0.76 0.82 0.79 850

average eval\_loss: 0.514488285667328, accuracy: 0.8247058823529412

training 4 epoch...

17.658214330673218 seconds used......

4 training finished! average train loss: 0.7472600395862873

total clf loss: 25.60719344764948 total adv loss: 11.482354598119855

evaluating...

Confusion matrix, without normalization

[[ 1 8 32]

[ 0 40 55]

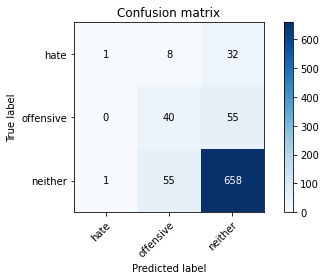
[ 1 55 658]]

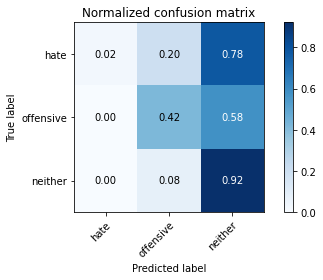
Normalized confusion matrix

[[0.02439024 0.19512195 0.7804878 ]

[0. 0.42105263 0.57894737]

[0.00140056 0.07703081 0.92156863]]





precision recall f1-score support

hate 0.50 0.02 0.05 41

offensive 0.39 0.42 0.40 95

neither 0.88 0.92 0.90 714

accuracy 0.82 850

macro avg 0.59 0.46 0.45 850

weighted avg 0.81 0.82 0.81 850

average eval\_loss: 0.4937313755237779, accuracy: 0.8223529411764706

training 5 epoch...

17.60097050666809 seconds used......

5 training finished! average train loss: 0.3647795524734717

total clf loss: 17.017090544104576 total adv loss: 3.346790240102564

evaluating...

Confusion matrix, without normalization

[[ 4 3 34]

[ 2 22 71]

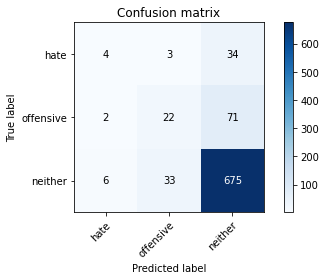
[ 6 33 675]]

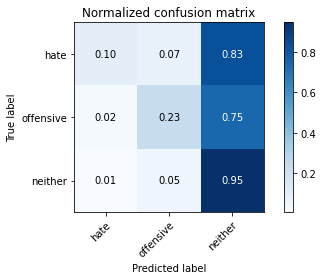
Normalized confusion matrix

[[0.09756098 0.07317073 0.82926829]

[0.02105263 0.23157895 0.74736842]

[0.00840336 0.04621849 0.94537815]]





precision recall f1-score support

hate 0.33 0.10 0.15 41

offensive 0.38 0.23 0.29 95

neither 0.87 0.95 0.90 714

accuracy 0.82 850

macro avg 0.53 0.42 0.45 850

weighted avg 0.79 0.82 0.80 850

average eval\_loss: 0.5960027006736328, accuracy: 0.8247058823529412

training 6 epoch...

17.692203760147095 seconds used......

6 training finished! average train loss: 0.21856732402856532

total clf loss: 12.290528770536184 total adv loss: 0.9581736286781961

evaluating...

Confusion matrix, without normalization

[[ 7 6 28]

[ 11 49 35]

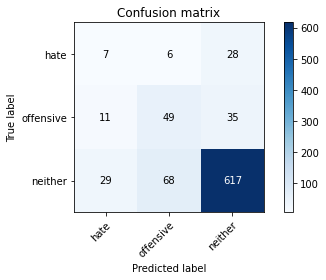
[ 29 68 617]]

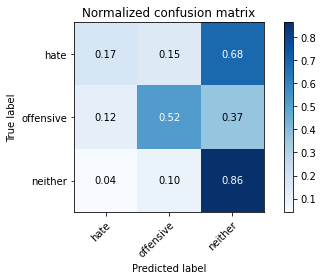
Normalized confusion matrix

[[0.17073171 0.14634146 0.68292683]

[0.11578947 0.51578947 0.36842105]

[0.04061625 0.0952381 0.86414566]]





precision recall f1-score support

hate 0.15 0.17 0.16 41

offensive 0.40 0.52 0.45 95

neither 0.91 0.86 0.89 714

accuracy 0.79 850

macro avg 0.48 0.52 0.50 850

weighted avg 0.81 0.79 0.80 850

average eval\_loss: 0.6113360181104305, accuracy: 0.7917647058823529

training 7 epoch...

17.642966747283936 seconds used......

7 training finished! average train loss: 0.1637633387859051

total clf loss: 9.314297918230295 total adv loss: 0.665159597782349

evaluating...

Confusion matrix, without normalization

[[ 10 0 31]

[ 6 10 79]

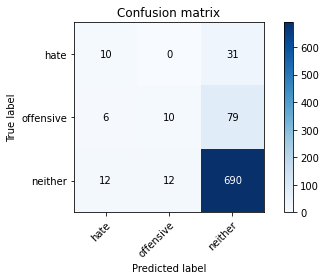
[ 12 12 690]]

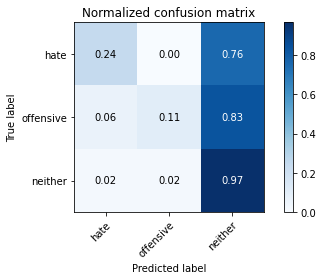
Normalized confusion matrix

[[0.24390244 0. 0.75609756]

[0.06315789 0.10526316 0.83157895]

[0.01680672 0.01680672 0.96638655]]





precision recall f1-score support

hate 0.36 0.24 0.29 41

offensive 0.45 0.11 0.17 95

neither 0.86 0.97 0.91 714

accuracy 0.84 850

macro avg 0.56 0.44 0.46 850

weighted avg 0.79 0.84 0.80 850

average eval\_loss: 0.7621278545905857, accuracy: 0.8352941176470589

training 8 epoch...

17.58217477798462 seconds used......

8 training finished! average train loss: 0.1513978112488985

total clf loss: 7.038230003789067 total adv loss: 1.4013138903028448

evaluating...

Confusion matrix, without normalization

[[ 7 1 33]

[ 6 15 74]

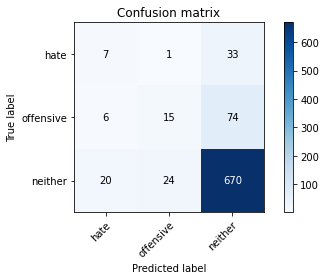
[ 20 24 670]]

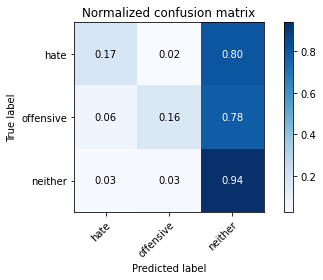
Normalized confusion matrix

[[0.17073171 0.02439024 0.80487805]

[0.06315789 0.15789474 0.77894737]

[0.0280112 0.03361345 0.93837535]]





precision recall f1-score support

hate 0.21 0.17 0.19 41

offensive 0.38 0.16 0.22 95

neither 0.86 0.94 0.90 714

accuracy 0.81 850

macro avg 0.48 0.42 0.44 850

weighted avg 0.78 0.81 0.79 850

average eval\_loss: 0.7382438152616974, accuracy: 0.8141176470588235

training 9 epoch...

17.688472270965576 seconds used......

9 training finished! average train loss: 0.10197356533115873

total clf loss: 5.244314060546458 total adv loss: 0.6919838803114544

evaluating...

Confusion matrix, without normalization

[[ 12 1 28]

[ 17 15 63]

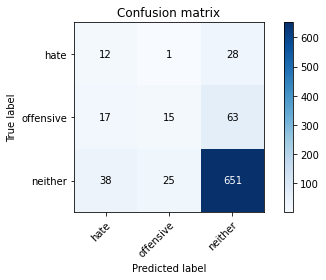
[ 38 25 651]]

Normalized confusion matrix

[[0.29268293 0.02439024 0.68292683]

[0.17894737 0.15789474 0.66315789]

[0.05322129 0.03501401 0.91176471]]





precision recall f1-score support

hate 0.18 0.29 0.22 41

offensive 0.37 0.16 0.22 95

neither 0.88 0.91 0.89 714

accuracy 0.80 850

macro avg 0.47 0.45 0.45 850

weighted avg 0.79 0.80 0.79 850

average eval\_loss: 0.7856083283482422, accuracy: 0.7976470588235294

training 10 epoch...

17.567150592803955 seconds used......

10 training finished! average train loss: 0.07286323925050406

total clf loss: 4.2001633788459 total adv loss: 0.26797358467865706

evaluating...

Confusion matrix, without normalization

[[ 13 3 25]

[ 17 25 53]

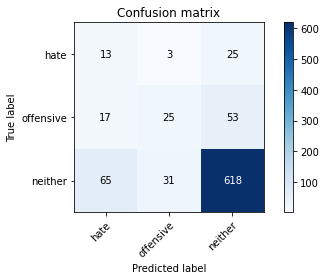
[ 65 31 618]]

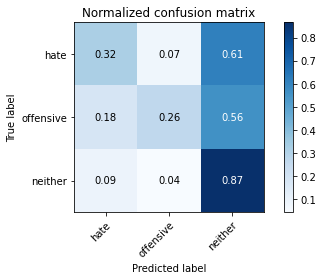
Normalized confusion matrix

[[0.31707317 0.07317073 0.6097561 ]

[0.17894737 0.26315789 0.55789474]

[0.09103641 0.04341737 0.86554622]]





precision recall f1-score support

hate 0.14 0.32 0.19 41

offensive 0.42 0.26 0.32 95

neither 0.89 0.87 0.88 714

accuracy 0.77 850

macro avg 0.48 0.48 0.46 850

weighted avg 0.80 0.77 0.78 850

average eval\_loss: 0.779861498856917, accuracy: 0.7717647058823529

4 run......

training 1 epoch...

17.671857833862305 seconds used......

1 training finished! average train loss: 1.8386301645865808

total clf loss: 52.28300577402115 total adv loss: 33.613977670669556

evaluating...

Confusion matrix, without normalization

[[ 0 0 41]

[ 0 0 95]

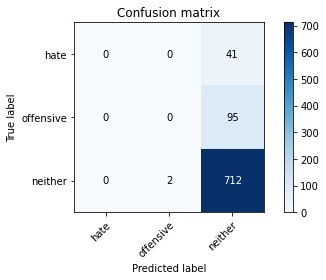
[ 0 2 712]]

Normalized confusion matrix

[[0. 0. 1. ]

[0. 0. 1. ]

[0. 0.00280112 0.99719888]]





/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.00 0.00 0.00 95

neither 0.84 1.00 0.91 714

accuracy 0.84 850

macro avg 0.28 0.33 0.30 850

weighted avg 0.71 0.84 0.77 850

average eval\_loss: 0.5949910842748576, accuracy: 0.8376470588235294

training 2 epoch...

17.692880392074585 seconds used......

2 training finished! average train loss: 1.5153901668695302

total clf loss: 44.31090211868286 total adv loss: 27.094729155302048

evaluating...

Confusion matrix, without normalization

[[ 0 9 32]

[ 0 24 71]

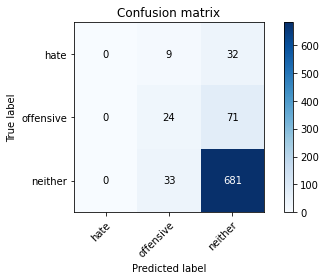
[ 0 33 681]]

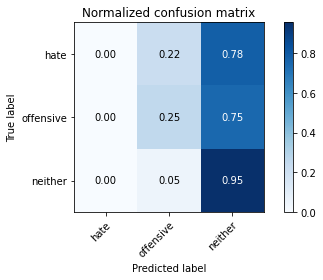
Normalized confusion matrix

[[0. 0.2195122 0.7804878 ]

[0. 0.25263158 0.74736842]

[0. 0.04621849 0.95378151]]





/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.36 0.25 0.30 95

neither 0.87 0.95 0.91 714

accuracy 0.83 850

macro avg 0.41 0.40 0.40 850

weighted avg 0.77 0.83 0.80 850

average eval\_loss: 0.5257994305602339, accuracy: 0.8294117647058824

training 3 epoch...

17.71853804588318 seconds used......

3 training finished! average train loss: 1.0927179088959327

total clf loss: 34.055740267038345 total adv loss: 18.485461801290512

evaluating...

Confusion matrix, without normalization

[[ 1 1 39]

[ 0 12 83]

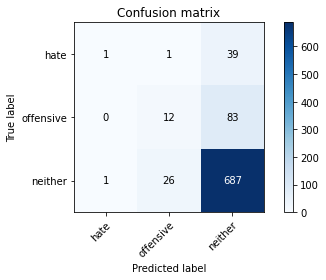
[ 1 26 687]]

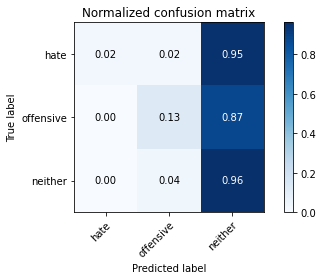
Normalized confusion matrix

[[0.02439024 0.02439024 0.95121951]

[0. 0.12631579 0.87368421]

[0.00140056 0.03641457 0.96218487]]





precision recall f1-score support

hate 0.50 0.02 0.05 41

offensive 0.31 0.13 0.18 95

neither 0.85 0.96 0.90 714

accuracy 0.82 850

macro avg 0.55 0.37 0.38 850

weighted avg 0.77 0.82 0.78 850

average eval\_loss: 0.5121309821775486, accuracy: 0.8235294117647058

training 4 epoch...

17.570088386535645 seconds used......

4 training finished! average train loss: 0.6863716531258363

total clf loss: 23.86722795665264 total adv loss: 10.37346478755353

evaluating...

Confusion matrix, without normalization

[[ 1 2 38]

[ 4 20 71]

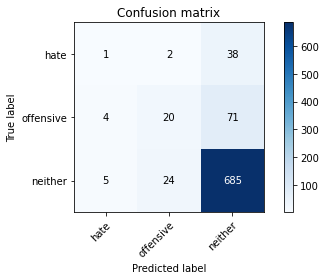
[ 5 24 685]]

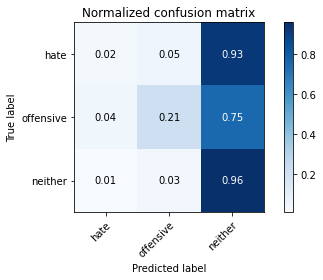
Normalized confusion matrix

[[0.02439024 0.04878049 0.92682927]

[0.04210526 0.21052632 0.74736842]

[0.0070028 0.03361345 0.95938375]]





precision recall f1-score support

hate 0.10 0.02 0.04 41

offensive 0.43 0.21 0.28 95

neither 0.86 0.96 0.91 714

accuracy 0.83 850

macro avg 0.47 0.40 0.41 850

weighted avg 0.78 0.83 0.80 850

average eval\_loss: 0.47817198256420534, accuracy: 0.8305882352941176

training 5 epoch...

17.70958423614502 seconds used......

5 training finished! average train loss: 0.3690680103806349

total clf loss: 16.517463766038418 total adv loss: 3.735978412485565

evaluating...

Confusion matrix, without normalization

[[ 2 6 33]

[ 3 44 48]

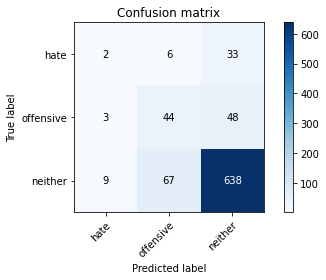
[ 9 67 638]]

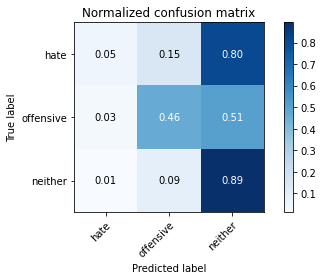
Normalized confusion matrix

[[0.04878049 0.14634146 0.80487805]

[0.03157895 0.46315789 0.50526316]

[0.01260504 0.09383754 0.89355742]]





precision recall f1-score support

hate 0.14 0.05 0.07 41

offensive 0.38 0.46 0.42 95

neither 0.89 0.89 0.89 714

accuracy 0.80 850

macro avg 0.47 0.47 0.46 850

weighted avg 0.79 0.80 0.80 850

average eval\_loss: 0.5508986054811367, accuracy: 0.8047058823529412

training 6 epoch...

17.636920928955078 seconds used......

6 training finished! average train loss: 0.19713921380730776

total clf loss: 11.076412700116634 total adv loss: 0.8688181014913425

evaluating...

Confusion matrix, without normalization

[[ 6 5 30]

[ 7 45 43]

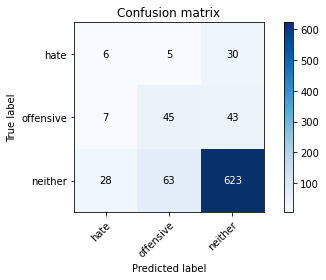
[ 28 63 623]]

Normalized confusion matrix

[[0.14634146 0.12195122 0.73170732]

[0.07368421 0.47368421 0.45263158]

[0.03921569 0.08823529 0.87254902]]





precision recall f1-score support

hate 0.15 0.15 0.15 41

offensive 0.40 0.47 0.43 95

neither 0.90 0.87 0.88 714

accuracy 0.79 850

macro avg 0.48 0.50 0.49 850

weighted avg 0.80 0.79 0.80 850

average eval\_loss: 0.6224932171361044, accuracy: 0.7929411764705883

training 7 epoch...

17.564486980438232 seconds used......

7 training finished! average train loss: 0.12533320893461888

total clf loss: 7.769125271588564 total adv loss: 0.18876663492119405

evaluating...

Confusion matrix, without normalization

[[ 11 12 18]

[ 13 55 27]

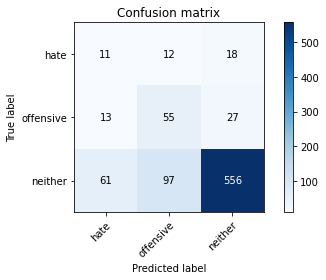
[ 61 97 556]]

Normalized confusion matrix

[[0.26829268 0.29268293 0.43902439]

[0.13684211 0.57894737 0.28421053]

[0.08543417 0.13585434 0.77871148]]





precision recall f1-score support

hate 0.13 0.27 0.17 41

offensive 0.34 0.58 0.42 95

neither 0.93 0.78 0.85 714

accuracy 0.73 850

macro avg 0.46 0.54 0.48 850

weighted avg 0.82 0.73 0.77 850

average eval\_loss: 0.8658456977045293, accuracy: 0.731764705882353

training 8 epoch...

17.666091442108154 seconds used......

8 training finished! average train loss: 0.10041964099957393

total clf loss: 5.640806877054274 total adv loss: 0.443234873129569

evaluating...

Confusion matrix, without normalization

[[ 13 6 22]

[ 20 39 36]

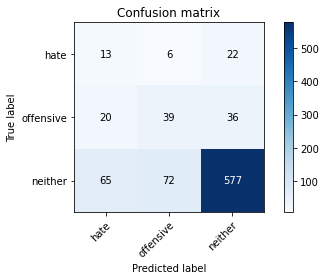
[ 65 72 577]]

Normalized confusion matrix

[[0.31707317 0.14634146 0.53658537]

[0.21052632 0.41052632 0.37894737]

[0.09103641 0.10084034 0.80812325]]





precision recall f1-score support

hate 0.13 0.32 0.19 41

offensive 0.33 0.41 0.37 95

neither 0.91 0.81 0.86 714

accuracy 0.74 850

macro avg 0.46 0.51 0.47 850

weighted avg 0.81 0.74 0.77 850

average eval\_loss: 0.8586019885964519, accuracy: 0.74

training 9 epoch...

17.58128547668457 seconds used......

9 training finished! average train loss: 0.18284890244786556

total clf loss: 5.863410614430904 total adv loss: 3.010884032693866

evaluating...

Confusion matrix, without normalization

[[ 14 4 23]

[ 10 23 62]

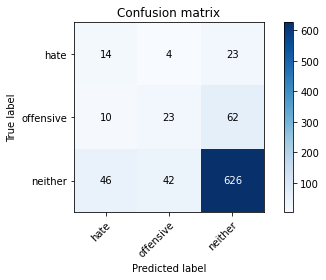
[ 46 42 626]]

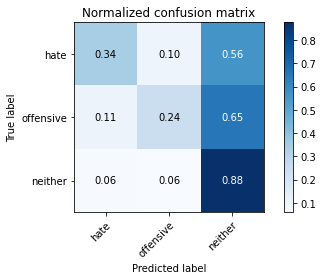
Normalized confusion matrix

[[0.34146341 0.09756098 0.56097561]

[0.10526316 0.24210526 0.65263158]

[0.06442577 0.05882353 0.8767507 ]]





precision recall f1-score support

hate 0.20 0.34 0.25 41

offensive 0.33 0.24 0.28 95

neither 0.88 0.88 0.88 714

accuracy 0.78 850

macro avg 0.47 0.49 0.47 850

weighted avg 0.79 0.78 0.78 850

average eval\_loss: 0.7459330993249665, accuracy: 0.78

training 10 epoch...

17.615699768066406 seconds used......

10 training finished! average train loss: 0.058375413586887026

total clf loss: 3.3625676166266203 total adv loss: 0.21591712866938906

evaluating...

Confusion matrix, without normalization

[[ 10 11 20]

[ 12 47 36]

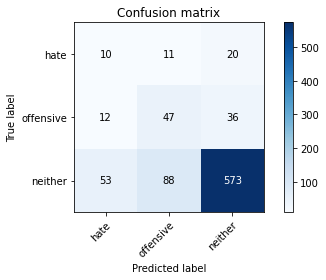
[ 53 88 573]]

Normalized confusion matrix

[[0.24390244 0.26829268 0.48780488]

[0.12631579 0.49473684 0.37894737]

[0.07422969 0.1232493 0.80252101]]





precision recall f1-score support

hate 0.13 0.24 0.17 41

offensive 0.32 0.49 0.39 95

neither 0.91 0.80 0.85 714

accuracy 0.74 850

macro avg 0.46 0.51 0.47 850

weighted avg 0.81 0.74 0.77 850

average eval\_loss: 0.8925791357178241, accuracy: 0.7411764705882353

5 run......

training 1 epoch...

17.694028854370117 seconds used......

1 training finished! average train loss: 1.8526459345450768

total clf loss: 51.94428777694702 total adv loss: 34.238848835229874

evaluating...

Confusion matrix, without normalization

[[ 0 0 41]

[ 0 0 95]

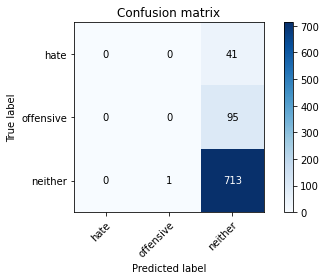
[ 0 1 713]]

Normalized confusion matrix

[[0. 0. 1. ]

[0. 0. 1. ]

[0. 0.00140056 0.99859944]]





/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.00 0.00 0.00 95

neither 0.84 1.00 0.91 714

accuracy 0.84 850

macro avg 0.28 0.33 0.30 850

weighted avg 0.71 0.84 0.77 850

average eval\_loss: 0.5867576578328776, accuracy: 0.8388235294117647

training 2 epoch...

17.792851448059082 seconds used......

2 training finished! average train loss: 1.5398565365717962

total clf loss: 44.253137677907944 total adv loss: 27.91876856982708

evaluating...

Confusion matrix, without normalization

[[ 0 12 29]

[ 0 35 60]

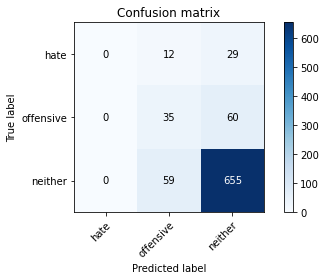
[ 0 59 655]]

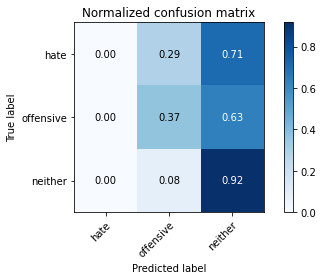
Normalized confusion matrix

[[0. 0.29268293 0.70731707]

[0. 0.36842105 0.63157895]

[0. 0.08263305 0.91736695]]





/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

/opt/conda/lib/python3.7/site-packages/sklearn/metrics/\_classification.py:1245: UndefinedMetricWarning: Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples. Use `zero\_division` parameter to control this behavior.

\_warn\_prf(average, modifier, msg\_start, len(result))

precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.33 0.37 0.35 95

neither 0.88 0.92 0.90 714

accuracy 0.81 850

macro avg 0.40 0.43 0.42 850

weighted avg 0.78 0.81 0.79 850

average eval\_loss: 0.5912431847564009, accuracy: 0.8117647058823529

training 3 epoch...

17.619064807891846 seconds used......

3 training finished! average train loss: 1.126694667339325

total clf loss: 34.05684891343117 total adv loss: 19.589152194559574

evaluating...

Confusion matrix, without normalization

[[ 0 9 32]

[ 0 36 59]

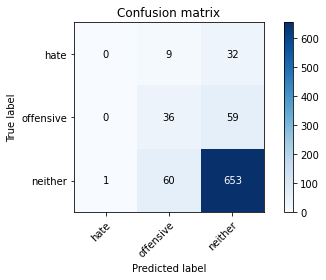
[ 1 60 653]]

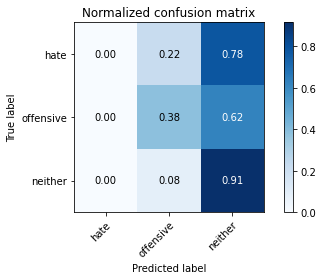
Normalized confusion matrix

[[0. 0.2195122 0.7804878 ]

[0. 0.37894737 0.62105263]

[0.00140056 0.08403361 0.91456583]]





precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.34 0.38 0.36 95

neither 0.88 0.91 0.90 714

accuracy 0.81 850

macro avg 0.41 0.43 0.42 850

weighted avg 0.78 0.81 0.79 850

average eval\_loss: 0.5092576878621827, accuracy: 0.8105882352941176

training 4 epoch...

17.700804710388184 seconds used......

4 training finished! average train loss: 0.6411149829626084

total clf loss: 22.56567084044218 total adv loss: 9.553401486948133

evaluating...

Confusion matrix, without normalization

[[ 2 14 25]

[ 4 43 48]

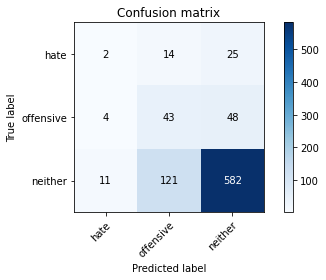
[ 11 121 582]]

Normalized confusion matrix

[[0.04878049 0.34146341 0.6097561 ]

[0.04210526 0.45263158 0.50526316]

[0.01540616 0.16946779 0.81512605]]





precision recall f1-score support

hate 0.12 0.05 0.07 41

offensive 0.24 0.45 0.32 95

neither 0.89 0.82 0.85 714

accuracy 0.74 850

macro avg 0.42 0.44 0.41 850

weighted avg 0.78 0.74 0.75 850

average eval\_loss: 0.6416272301833297, accuracy: 0.7376470588235294

training 5 epoch...

17.667529821395874 seconds used......

5 training finished! average train loss: 0.3109262647537085

total clf loss: 14.121839568018913 total adv loss: 3.0441838041879237

evaluating...

Confusion matrix, without normalization

[[ 8 4 29]

[ 5 22 68]

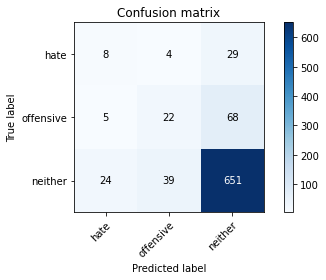
[ 24 39 651]]

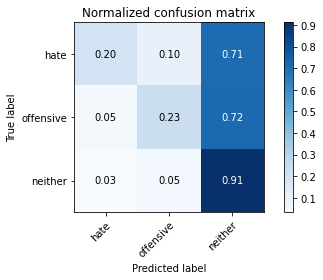
Normalized confusion matrix

[[0.19512195 0.09756098 0.70731707]

[0.05263158 0.23157895 0.71578947]

[0.03361345 0.05462185 0.91176471]]





precision recall f1-score support

hate 0.22 0.20 0.21 41

offensive 0.34 0.23 0.28 95

neither 0.87 0.91 0.89 714

accuracy 0.80 850

macro avg 0.47 0.45 0.46 850

weighted avg 0.78 0.80 0.79 850

average eval\_loss: 0.592383366294725, accuracy: 0.8011764705882353

training 6 epoch...

17.543388843536377 seconds used......

6 training finished! average train loss: 0.1812489914206358

total clf loss: 9.601088643074036 total adv loss: 1.0900478942057816

evaluating...

Confusion matrix, without normalization

[[ 13 3 25]

[ 12 32 51]

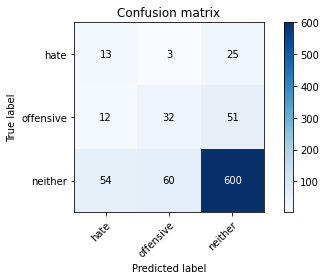
[ 54 60 600]]

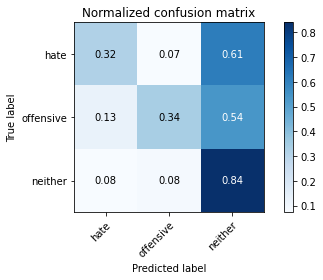
Normalized confusion matrix

[[0.31707317 0.07317073 0.6097561 ]

[0.12631579 0.33684211 0.53684211]

[0.07563025 0.08403361 0.84033613]]





precision recall f1-score support

hate 0.16 0.32 0.22 41

offensive 0.34 0.34 0.34 95

neither 0.89 0.84 0.86 714

accuracy 0.76 850

macro avg 0.46 0.50 0.47 850

weighted avg 0.79 0.76 0.77 850

average eval\_loss: 0.6971082209154617, accuracy: 0.7588235294117647

training 7 epoch...

17.679246425628662 seconds used......

7 training finished! average train loss: 0.18578347735680067

total clf loss: 8.161612540483475 total adv loss: 1.9571567877719644

evaluating...

Confusion matrix, without normalization

[[ 13 5 23]

[ 7 12 76]

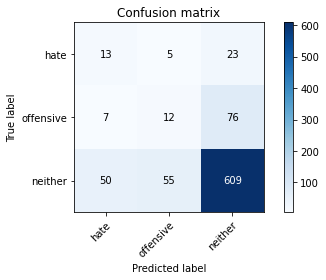
[ 50 55 609]]

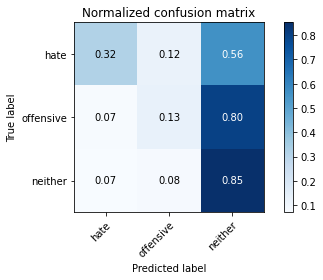
Normalized confusion matrix

[[0.31707317 0.12195122 0.56097561]

[0.07368421 0.12631579 0.8 ]

[0.07002801 0.07703081 0.85294118]]





precision recall f1-score support

hate 0.19 0.32 0.23 41

offensive 0.17 0.13 0.14 95

neither 0.86 0.85 0.86 714

accuracy 0.75 850

macro avg 0.40 0.43 0.41 850

weighted avg 0.75 0.75 0.75 850

average eval\_loss: 0.8727161321832344, accuracy: 0.7458823529411764

training 8 epoch...

17.65022587776184 seconds used......

8 training finished! average train loss: 0.08905860058103617

total clf loss: 5.039771712385118 total adv loss: 0.3745186585583724

evaluating...

Confusion matrix, without normalization

[[ 8 4 29]

[ 6 34 55]

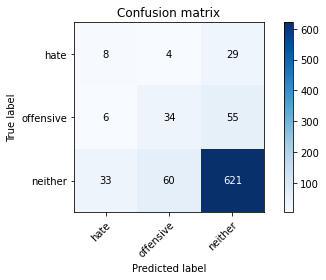
[ 33 60 621]]

Normalized confusion matrix

[[0.19512195 0.09756098 0.70731707]

[0.06315789 0.35789474 0.57894737]

[0.04621849 0.08403361 0.8697479 ]]





precision recall f1-score support

hate 0.17 0.20 0.18 41

offensive 0.35 0.36 0.35 95

neither 0.88 0.87 0.88 714

accuracy 0.78 850

macro avg 0.47 0.47 0.47 850

weighted avg 0.79 0.78 0.78 850

average eval\_loss: 0.7418361505612644, accuracy: 0.78

training 9 epoch...

17.551318407058716 seconds used......

9 training finished! average train loss: 0.04693718182925995

total clf loss: 2.9858722696080804 total adv loss: 0.03252228652308986

evaluating...

Confusion matrix, without normalization

[[ 11 5 25]

[ 8 32 55]

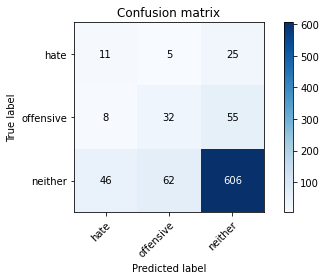
[ 46 62 606]]

Normalized confusion matrix

[[0.26829268 0.12195122 0.6097561 ]

[0.08421053 0.33684211 0.57894737]

[0.06442577 0.08683473 0.8487395 ]]





precision recall f1-score support

hate 0.17 0.27 0.21 41

offensive 0.32 0.34 0.33 95

neither 0.88 0.85 0.87 714

accuracy 0.76 850

macro avg 0.46 0.48 0.47 850

weighted avg 0.79 0.76 0.77 850

average eval\_loss: 0.8203508734925066, accuracy: 0.7635294117647059

training 10 epoch...

17.642605304718018 seconds used......

10 training finished! average train loss: 0.04134974355021348

total clf loss: 1.9762950658332556 total adv loss: 0.35571913781905096

evaluating...

Confusion matrix, without normalization

[[ 18 4 19]

[ 15 31 49]

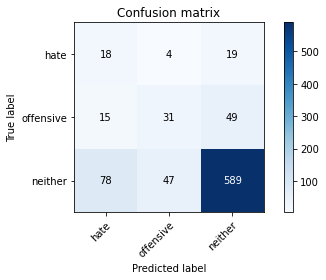
[ 78 47 589]]

Normalized confusion matrix

[[0.43902439 0.09756098 0.46341463]

[0.15789474 0.32631579 0.51578947]

[0.1092437 0.06582633 0.82492997]]





precision recall f1-score support

hate 0.16 0.44 0.24 41

offensive 0.38 0.33 0.35 95

neither 0.90 0.82 0.86 714

accuracy 0.75 850

macro avg 0.48 0.53 0.48 850

weighted avg 0.80 0.75 0.77 850

average eval\_loss: 0.9100192250889662, accuracy: 0.7505882352941177