1 run......

Downloading: 100%

464/464 [00:00<00:00, 10.9kB/s]

Downloading: 100%

270M/270M [00:10<00:00, 28.5MB/s]

training 1 epoch...

10.503941059112549 seconds used......

1 training finished! average train loss: 2.4287920594215393

total clf loss: 39.20946794748306 total adv loss: 24.1135233938694

evaluating...

Confusion matrix, without normalization

[[ 0 25 16]

[ 0 11 84]

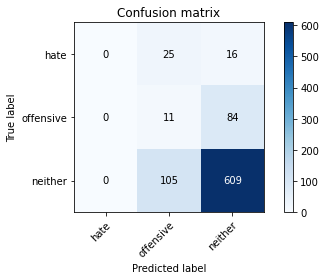
[ 0 105 609]]

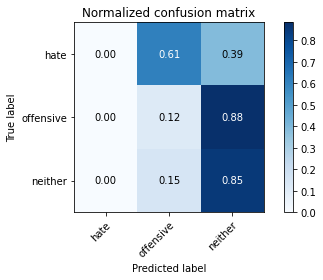
Normalized confusion matrix

[[0. 0.6097561 0.3902439 ]

[0. 0.11578947 0.88421053]

[0. 0.14705882 0.85294118]]





/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 0.08 0.12 0.09 95

neither 0.86 0.85 0.86 714

accuracy 0.73 850

macro avg 0.31 0.32 0.32 850

weighted avg 0.73 0.73 0.73 850

average eval\_loss: 0.8123534730700559, accuracy: 0.7294117647058823

training 2 epoch...

9.387865781784058 seconds used......

2 training finished! average train loss: 2.133102234866884

total clf loss: 34.51063400506973 total adv loss: 21.140522867441177

evaluating...

Confusion matrix, without normalization

[[ 0 35 6]

[ 0 82 13]

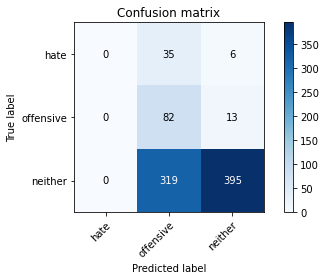
[ 0 319 395]]

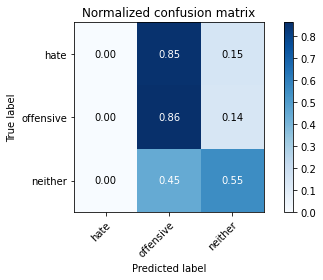
Normalized confusion matrix

[[0. 0.85365854 0.14634146]

[0. 0.86315789 0.13684211]

[0. 0.44677871 0.55322129]]





/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 0.19 0.86 0.31 95

neither 0.95 0.55 0.70 714

accuracy 0.56 850

macro avg 0.38 0.47 0.34 850

weighted avg 0.82 0.56 0.62 850

average eval\_loss: 0.861003186467082, accuracy: 0.5611764705882353

training 3 epoch...

9.328167915344238 seconds used......

3 training finished! average train loss: 1.7780425316757626

total clf loss: 29.84712427854538 total adv loss: 17.08120322227478

evaluating...

Confusion matrix, without normalization

[[ 0 31 10]

[ 0 73 22]

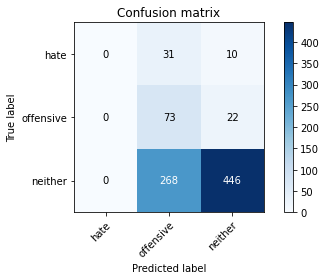
[ 0 268 446]]

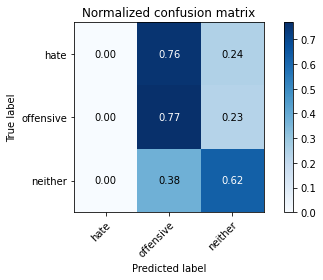
Normalized confusion matrix

[[0. 0.75609756 0.24390244]

[0. 0.76842105 0.23157895]

[0. 0.37535014 0.62464986]]





/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 0.20 0.77 0.31 95

neither 0.93 0.62 0.75 714

accuracy 0.61 850

macro avg 0.38 0.46 0.35 850

weighted avg 0.81 0.61 0.66 850

average eval\_loss: 0.8152948143177254, accuracy: 0.6105882352941177

training 4 epoch...

9.410498857498169 seconds used......

4 training finished! average train loss: 1.1845377170377307

total clf loss: 23.09373649954796 total adv loss: 9.774810690432787

evaluating...

Confusion matrix, without normalization

[[ 0 21 20]

[ 0 55 40]

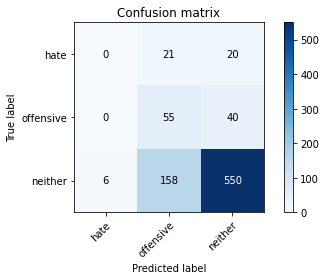
[ 6 158 550]]

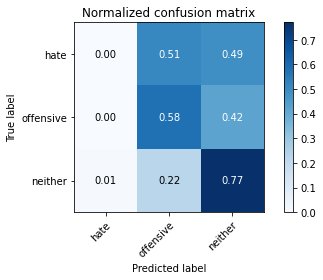
Normalized confusion matrix

[[0. 0.51219512 0.48780488]

[0. 0.57894737 0.42105263]

[0.00840336 0.22128852 0.77030812]]





precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.24 0.58 0.33 95

neither 0.90 0.77 0.83 714

accuracy 0.71 850

macro avg 0.38 0.45 0.39 850

weighted avg 0.78 0.71 0.74 850

average eval\_loss: 0.7182138396730257, accuracy: 0.711764705882353

training 5 epoch...

9.348937034606934 seconds used......

5 training finished! average train loss: 0.7111764786144098

total clf loss: 16.506112091243267 total adv loss: 4.548120545689017

evaluating...

Confusion matrix, without normalization

[[ 6 20 15]

[ 4 67 24]

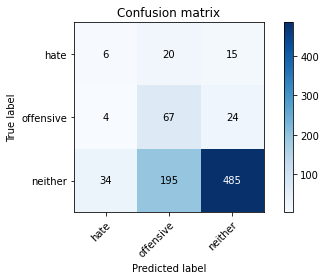
[ 34 195 485]]

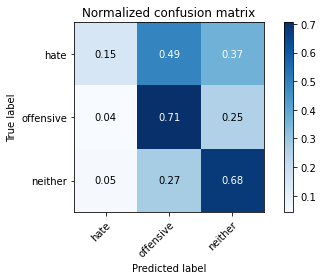
Normalized confusion matrix

[[0.14634146 0.48780488 0.36585366]

[0.04210526 0.70526316 0.25263158]

[0.04761905 0.27310924 0.67927171]]





precision recall f1-score support

hate 0.14 0.15 0.14 41

offensive 0.24 0.71 0.36 95

neither 0.93 0.68 0.78 714

accuracy 0.66 850

macro avg 0.43 0.51 0.43 850

weighted avg 0.81 0.66 0.70 850

average eval\_loss: 0.8703128305632014, accuracy: 0.6564705882352941

training 6 epoch...

9.429388999938965 seconds used......

6 training finished! average train loss: 0.3472372976442178

total clf loss: 10.856822397559881 total adv loss: 0.8218601408880204

evaluating...

Confusion matrix, without normalization

[[ 8 20 13]

[ 6 78 11]

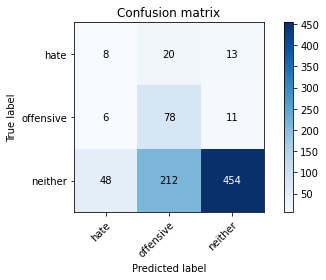
[ 48 212 454]]

Normalized confusion matrix

[[0.19512195 0.48780488 0.31707317]

[0.06315789 0.82105263 0.11578947]

[0.06722689 0.29691877 0.63585434]]





precision recall f1-score support

hate 0.13 0.20 0.16 41

offensive 0.25 0.82 0.39 95

neither 0.95 0.64 0.76 714

accuracy 0.64 850

macro avg 0.44 0.55 0.43 850

weighted avg 0.83 0.64 0.69 850

average eval\_loss: 0.9405198784675016, accuracy: 0.6352941176470588

training 7 epoch...

9.384620666503906 seconds used......

7 training finished! average train loss: 0.22366728405985567

total clf loss: 7.787458598613739 total adv loss: 0.13228180301666725

evaluating...

Confusion matrix, without normalization

[[ 8 16 17]

[ 9 63 23]

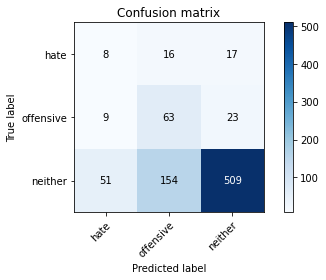
[ 51 154 509]]

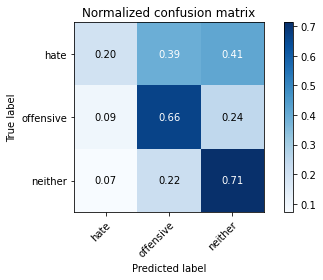
Normalized confusion matrix

[[0.19512195 0.3902439 0.41463415]

[0.09473684 0.66315789 0.24210526]

[0.07142857 0.21568627 0.71288515]]





precision recall f1-score support

hate 0.12 0.20 0.15 41

offensive 0.27 0.66 0.38 95

neither 0.93 0.71 0.81 714

accuracy 0.68 850

macro avg 0.44 0.52 0.45 850

weighted avg 0.81 0.68 0.73 850

average eval\_loss: 0.8723462354391813, accuracy: 0.6823529411764706

training 8 epoch...

9.441221952438354 seconds used......

8 training finished! average train loss: 0.1710649408503539

total clf loss: 6.017715675756335 total adv loss: 0.07031109421950532

evaluating...

Confusion matrix, without normalization

[[ 8 22 11]

[ 8 76 11]

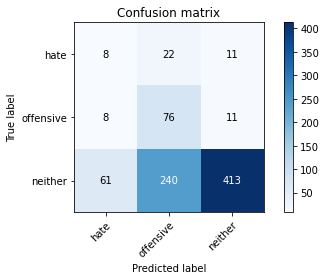
[ 61 240 413]]

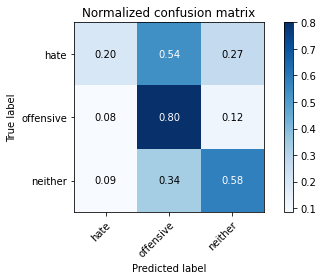
Normalized confusion matrix

[[0.19512195 0.53658537 0.26829268]

[0.08421053 0.8 0.11578947]

[0.08543417 0.33613445 0.57843137]]





precision recall f1-score support

hate 0.10 0.20 0.14 41

offensive 0.22 0.80 0.35 95

neither 0.95 0.58 0.72 714

accuracy 0.58 850

macro avg 0.43 0.52 0.40 850

weighted avg 0.83 0.58 0.65 850

average eval\_loss: 1.3222788035609694, accuracy: 0.5847058823529412

training 9 epoch...

9.43401312828064 seconds used......

9 training finished! average train loss: 0.23473019525408745

total clf loss: 5.286435227841139 total adv loss: 1.5819258545379853

evaluating...

Confusion matrix, without normalization

[[ 12 15 14]

[ 11 59 25]

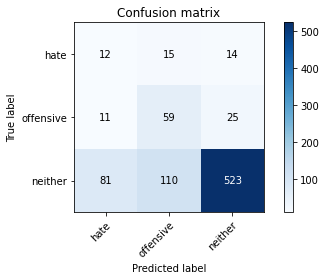
[ 81 110 523]]

Normalized confusion matrix

[[0.29268293 0.36585366 0.34146341]

[0.11578947 0.62105263 0.26315789]

[0.11344538 0.15406162 0.732493 ]]





precision recall f1-score support

hate 0.12 0.29 0.17 41

offensive 0.32 0.62 0.42 95

neither 0.93 0.73 0.82 714

accuracy 0.70 850

macro avg 0.46 0.55 0.47 850

weighted avg 0.82 0.70 0.74 850

average eval\_loss: 0.8352131578898014, accuracy: 0.6988235294117647

training 10 epoch...

9.447288990020752 seconds used......

10 training finished! average train loss: 0.13148933302404153

total clf loss: 4.173304103314877 total adv loss: 0.28015594221142237

evaluating...

Confusion matrix, without normalization

[[ 14 13 14]

[ 17 57 21]

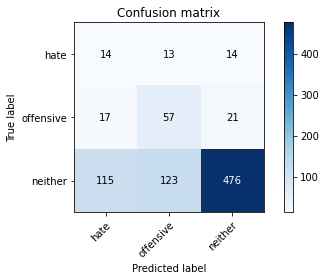
[115 123 476]]

Normalized confusion matrix

[[0.34146341 0.31707317 0.34146341]

[0.17894737 0.6 0.22105263]

[0.16106443 0.17226891 0.66666667]]





precision recall f1-score support

hate 0.10 0.34 0.15 41

offensive 0.30 0.60 0.40 95

neither 0.93 0.67 0.78 714

accuracy 0.64 850

macro avg 0.44 0.54 0.44 850

weighted avg 0.82 0.64 0.70 850

average eval\_loss: 1.044161303921841, accuracy: 0.6435294117647059

2 run......

training 1 epoch...

9.454328775405884 seconds used......

1 training finished! average train loss: 2.451566603448656

total clf loss: 39.018264055252075 total adv loss: 24.61906659603119

evaluating...

Confusion matrix, without normalization

[[ 0 34 7]

[ 0 42 53]

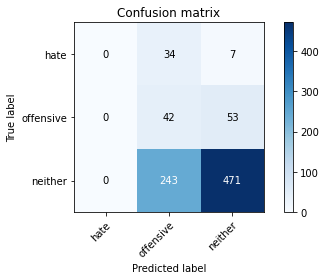
[ 0 243 471]]

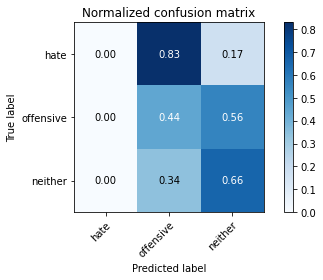
Normalized confusion matrix

[[0. 0.82926829 0.17073171]

[0. 0.44210526 0.55789474]

[0. 0.34033613 0.65966387]]





/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 0.13 0.44 0.20 95

neither 0.89 0.66 0.76 714

accuracy 0.60 850

macro avg 0.34 0.37 0.32 850

weighted avg 0.76 0.60 0.66 850

average eval\_loss: 0.9004825866499613, accuracy: 0.6035294117647059

training 2 epoch...

9.365827798843384 seconds used......

2 training finished! average train loss: 2.1565990911589727

total clf loss: 34.56757414340973 total adv loss: 21.534996390342712

evaluating...

Confusion matrix, without normalization

[[ 0 25 16]

[ 0 44 51]

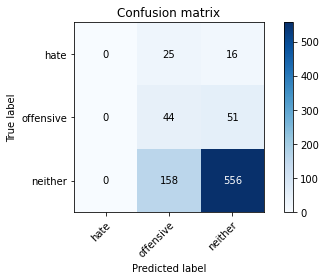
[ 0 158 556]]

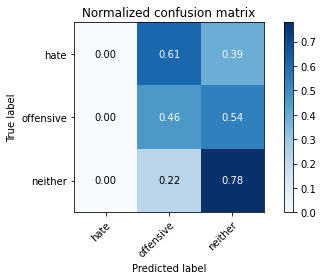
Normalized confusion matrix

[[0. 0.6097561 0.3902439 ]

[0. 0.46315789 0.53684211]

[0. 0.22128852 0.77871148]]





/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 0.19 0.46 0.27 95

neither 0.89 0.78 0.83 714

accuracy 0.71 850

macro avg 0.36 0.41 0.37 850

weighted avg 0.77 0.71 0.73 850

average eval\_loss: 0.7861052450052527, accuracy: 0.7058823529411765

training 3 epoch...

9.450908422470093 seconds used......

3 training finished! average train loss: 1.751023679971695

total clf loss: 29.666287004947662 total adv loss: 16.68528261780739

evaluating...

Confusion matrix, without normalization

[[ 2 29 10]

[ 0 80 15]

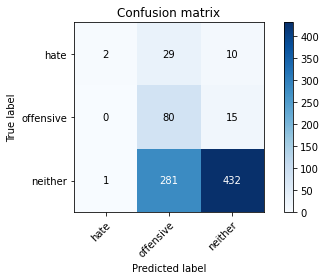
[ 1 281 432]]

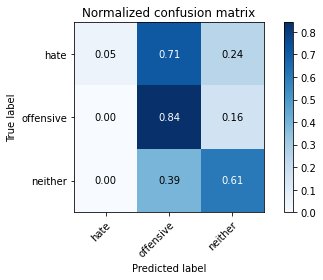
Normalized confusion matrix

[[0.04878049 0.70731707 0.24390244]

[0. 0.84210526 0.15789474]

[0.00140056 0.39355742 0.60504202]]





precision recall f1-score support

hate 0.67 0.05 0.09 41

offensive 0.21 0.84 0.33 95

neither 0.95 0.61 0.74 714

accuracy 0.60 850

macro avg 0.61 0.50 0.39 850

weighted avg 0.85 0.60 0.66 850

average eval\_loss: 0.8747103190006211, accuracy: 0.6047058823529412

training 4 epoch...

9.433327198028564 seconds used......

4 training finished! average train loss: 1.2149096503853798

total clf loss: 23.047888457775116 total adv loss: 10.344429474323988

evaluating...

Confusion matrix, without normalization

[[ 3 34 4]

[ 3 84 8]

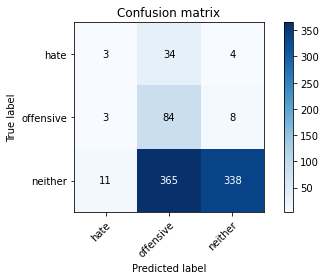
[ 11 365 338]]

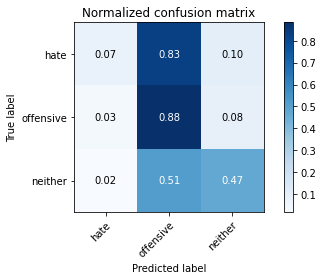
Normalized confusion matrix

[[0.07317073 0.82926829 0.09756098]

[0.03157895 0.88421053 0.08421053]

[0.01540616 0.51120448 0.47338936]]





precision recall f1-score support

hate 0.18 0.07 0.10 41

offensive 0.17 0.88 0.29 95

neither 0.97 0.47 0.64 714

accuracy 0.50 850

macro avg 0.44 0.48 0.34 850

weighted avg 0.84 0.50 0.57 850

average eval\_loss: 1.116603983176309, accuracy: 0.5

training 5 epoch...

9.358214378356934 seconds used......

5 training finished! average train loss: 0.7047993333803283

total clf loss: 16.872210681438446 total adv loss: 4.250282833119854

evaluating...

Confusion matrix, without normalization

[[ 14 14 13]

[ 10 57 28]

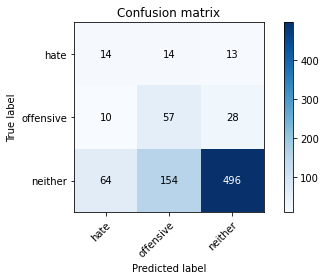
[ 64 154 496]]

Normalized confusion matrix

[[0.34146341 0.34146341 0.31707317]

[0.10526316 0.6 0.29473684]

[0.08963585 0.21568627 0.69467787]]





precision recall f1-score support

hate 0.16 0.34 0.22 41

offensive 0.25 0.60 0.36 95

neither 0.92 0.69 0.79 714

accuracy 0.67 850

macro avg 0.45 0.55 0.46 850

weighted avg 0.81 0.67 0.72 850

average eval\_loss: 0.8348794182719186, accuracy: 0.6670588235294118

training 6 epoch...

9.364153861999512 seconds used......

6 training finished! average train loss: 0.37252358595530194

total clf loss: 10.852833785116673 total adv loss: 1.2790076327510178

evaluating...

Confusion matrix, without normalization

[[ 11 17 13]

[ 9 52 34]

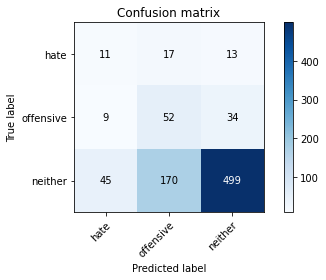
[ 45 170 499]]

Normalized confusion matrix

[[0.26829268 0.41463415 0.31707317]

[0.09473684 0.54736842 0.35789474]

[0.06302521 0.23809524 0.69887955]]





precision recall f1-score support

hate 0.17 0.27 0.21 41

offensive 0.22 0.55 0.31 95

neither 0.91 0.70 0.79 714

accuracy 0.66 850

macro avg 0.43 0.50 0.44 850

weighted avg 0.80 0.66 0.71 850

average eval\_loss: 0.8758790979898253, accuracy: 0.6611764705882353

training 7 epoch...

9.398285627365112 seconds used......

7 training finished! average train loss: 0.21652231727623278

total clf loss: 7.6345782075077295 total adv loss: 0.08011260248895269

evaluating...

Confusion matrix, without normalization

[[ 12 16 13]

[ 8 60 27]

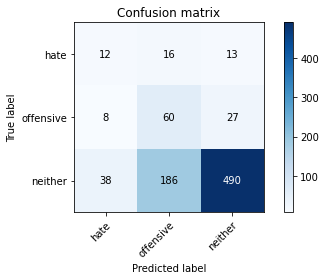
[ 38 186 490]]

Normalized confusion matrix

[[0.29268293 0.3902439 0.31707317]

[0.08421053 0.63157895 0.28421053]

[0.05322129 0.2605042 0.68627451]]





precision recall f1-score support

hate 0.21 0.29 0.24 41

offensive 0.23 0.63 0.34 95

neither 0.92 0.69 0.79 714

accuracy 0.66 850

macro avg 0.45 0.54 0.46 850

weighted avg 0.81 0.66 0.71 850

average eval\_loss: 0.9513975973565911, accuracy: 0.6611764705882353

training 8 epoch...

9.352762222290039 seconds used......

8 training finished! average train loss: 0.1659573253451122

total clf loss: 5.822494179010391 total adv loss: 0.07598478824365884

evaluating...

Confusion matrix, without normalization

[[ 11 17 13]

[ 7 65 23]

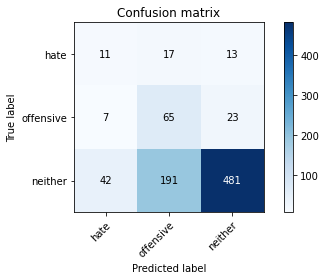
[ 42 191 481]]

Normalized confusion matrix

[[0.26829268 0.41463415 0.31707317]

[0.07368421 0.68421053 0.24210526]

[0.05882353 0.267507 0.67366947]]





precision recall f1-score support

hate 0.18 0.27 0.22 41

offensive 0.24 0.68 0.35 95

neither 0.93 0.67 0.78 714

accuracy 0.66 850

macro avg 0.45 0.54 0.45 850

weighted avg 0.82 0.66 0.71 850

average eval\_loss: 1.07084724748897, accuracy: 0.6552941176470588

training 9 epoch...

9.384487390518188 seconds used......

9 training finished! average train loss: 0.1936630855521394

total clf loss: 4.755922058597207 total adv loss: 1.1079745057395485

evaluating...

Confusion matrix, without normalization

[[ 23 5 13]

[ 27 29 39]

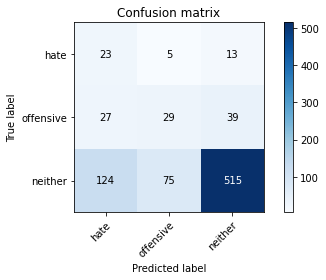
[124 75 515]]

Normalized confusion matrix

[[0.56097561 0.12195122 0.31707317]

[0.28421053 0.30526316 0.41052632]

[0.17366947 0.10504202 0.72128852]]





precision recall f1-score support

hate 0.13 0.56 0.21 41

offensive 0.27 0.31 0.28 95

neither 0.91 0.72 0.80 714

accuracy 0.67 850

macro avg 0.44 0.53 0.43 850

weighted avg 0.80 0.67 0.72 850

average eval\_loss: 0.9598384320216123, accuracy: 0.6670588235294118

training 10 epoch...

9.337589740753174 seconds used......

10 training finished! average train loss: 0.20504437169680992

total clf loss: 4.901811897754669 total adv loss: 1.239892747340491

evaluating...

Confusion matrix, without normalization

[[ 21 8 12]

[ 18 40 37]

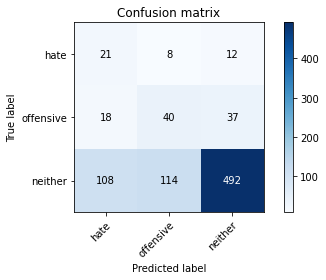
[108 114 492]]

Normalized confusion matrix

[[0.51219512 0.19512195 0.29268293]

[0.18947368 0.42105263 0.38947368]

[0.1512605 0.15966387 0.68907563]]





precision recall f1-score support

hate 0.14 0.51 0.22 41

offensive 0.25 0.42 0.31 95

neither 0.91 0.69 0.78 714

accuracy 0.65 850

macro avg 0.43 0.54 0.44 850

weighted avg 0.80 0.65 0.70 850

average eval\_loss: 1.0097995636941388, accuracy: 0.6505882352941177

3 run......

training 1 epoch...

9.395314455032349 seconds used......

1 training finished! average train loss: 2.3643946846326194

total clf loss: 37.50899064540863 total adv loss: 23.80460923910141

evaluating...

Confusion matrix, without normalization

[[ 0 36 5]

[ 0 60 35]

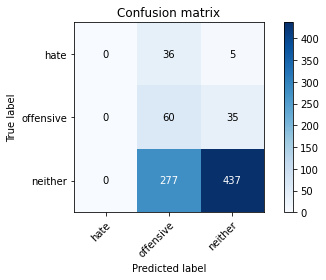
[ 0 277 437]]

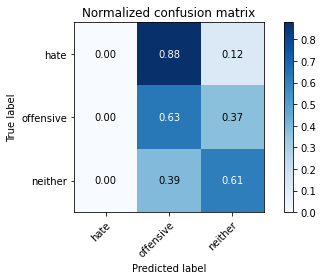
Normalized confusion matrix

[[0. 0.87804878 0.12195122]

[0. 0.63157895 0.36842105]

[0. 0.38795518 0.61204482]]





/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 0.16 0.63 0.26 95

neither 0.92 0.61 0.73 714

accuracy 0.58 850

macro avg 0.36 0.41 0.33 850

weighted avg 0.79 0.58 0.65 850

average eval\_loss: 0.8569963338763215, accuracy: 0.5847058823529412

training 2 epoch...

9.3530855178833 seconds used......

2 training finished! average train loss: 2.0231097075674267

total clf loss: 32.31685811281204 total adv loss: 20.257545560598373

evaluating...

Confusion matrix, without normalization

[[ 2 30 9]

[ 0 66 29]

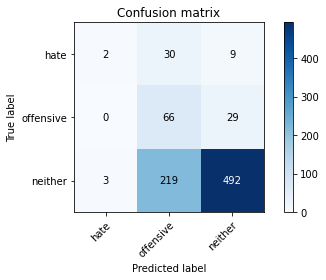
[ 3 219 492]]

Normalized confusion matrix

[[0.04878049 0.73170732 0.2195122 ]

[0. 0.69473684 0.30526316]

[0.00420168 0.30672269 0.68907563]]





precision recall f1-score support

hate 0.40 0.05 0.09 41

offensive 0.21 0.69 0.32 95

neither 0.93 0.69 0.79 714

accuracy 0.66 850

macro avg 0.51 0.48 0.40 850

weighted avg 0.82 0.66 0.70 850

average eval\_loss: 0.8245387877835784, accuracy: 0.6588235294117647

training 3 epoch...

9.409434795379639 seconds used......

3 training finished! average train loss: 1.6578859123918745

total clf loss: 27.173287630081177 total adv loss: 16.255302608013153

evaluating...

Confusion matrix, without normalization

[[ 5 28 8]

[ 5 73 17]

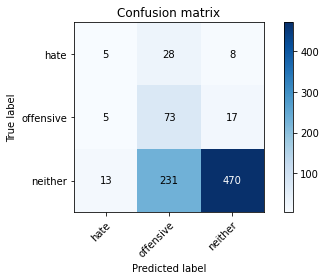
[ 13 231 470]]

Normalized confusion matrix

[[0.12195122 0.68292683 0.19512195]

[0.05263158 0.76842105 0.17894737]

[0.01820728 0.32352941 0.65826331]]





precision recall f1-score support

hate 0.22 0.12 0.16 41

offensive 0.22 0.77 0.34 95

neither 0.95 0.66 0.78 714

accuracy 0.64 850

macro avg 0.46 0.52 0.43 850

weighted avg 0.83 0.64 0.70 850

average eval\_loss: 0.8201872815572938, accuracy: 0.6447058823529411

training 4 epoch...

9.356668710708618 seconds used......

4 training finished! average train loss: 1.1643692023224301

total clf loss: 20.2029609978199 total adv loss: 10.857165090739727

evaluating...

Confusion matrix, without normalization

[[ 10 22 9]

[ 16 60 19]

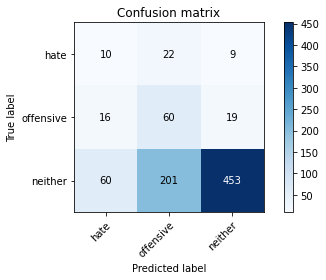
[ 60 201 453]]

Normalized confusion matrix

[[0.24390244 0.53658537 0.2195122 ]

[0.16842105 0.63157895 0.2 ]

[0.08403361 0.28151261 0.63445378]]





precision recall f1-score support

hate 0.12 0.24 0.16 41

offensive 0.21 0.63 0.32 95

neither 0.94 0.63 0.76 714

accuracy 0.62 850

macro avg 0.42 0.50 0.41 850

weighted avg 0.82 0.62 0.68 850

average eval\_loss: 0.8739289752272672, accuracy: 0.6152941176470588

training 5 epoch...

9.420047521591187 seconds used......

5 training finished! average train loss: 0.6965365587837167

total clf loss: 14.219030603766441 total adv loss: 5.428142710588872

evaluating...

Confusion matrix, without normalization

[[ 8 25 8]

[ 7 82 6]

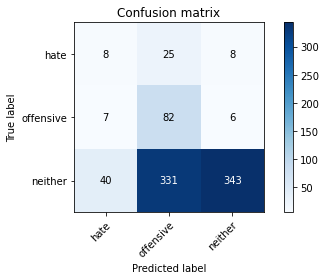
[ 40 331 343]]

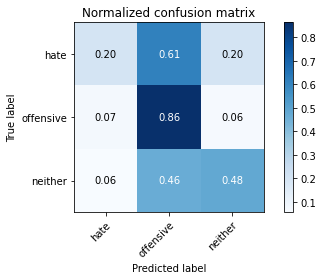
Normalized confusion matrix

[[0.19512195 0.6097561 0.19512195]

[0.07368421 0.86315789 0.06315789]

[0.05602241 0.46358543 0.48039216]]





precision recall f1-score support

hate 0.15 0.20 0.17 41

offensive 0.19 0.86 0.31 95

neither 0.96 0.48 0.64 714

accuracy 0.51 850

macro avg 0.43 0.51 0.37 850

weighted avg 0.83 0.51 0.58 850

average eval\_loss: 1.224537502887637, accuracy: 0.5094117647058823

training 6 epoch...

9.33538007736206 seconds used......

6 training finished! average train loss: 0.3964639880384008

total clf loss: 9.962823126465082 total adv loss: 2.1549402622622438

evaluating...

Confusion matrix, without normalization

[[ 9 15 17]

[ 12 50 33]

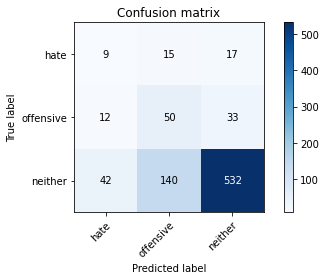
[ 42 140 532]]

Normalized confusion matrix

[[0.2195122 0.36585366 0.41463415]

[0.12631579 0.52631579 0.34736842]

[0.05882353 0.19607843 0.74509804]]





precision recall f1-score support

hate 0.14 0.22 0.17 41

offensive 0.24 0.53 0.33 95

neither 0.91 0.75 0.82 714

accuracy 0.70 850

macro avg 0.43 0.50 0.44 850

weighted avg 0.80 0.70 0.74 850

average eval\_loss: 0.7683047654843608, accuracy: 0.6952941176470588

training 7 epoch...

9.372700691223145 seconds used......

7 training finished! average train loss: 0.21269971277150843

total clf loss: 6.873874252662063 total adv loss: 0.3916577232466807

evaluating...

Confusion matrix, without normalization

[[ 13 18 10]

[ 14 58 23]

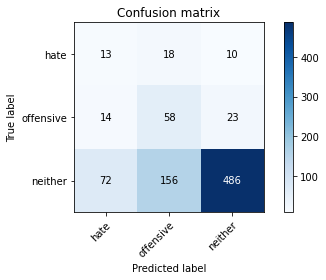
[ 72 156 486]]

Normalized confusion matrix

[[0.31707317 0.43902439 0.24390244]

[0.14736842 0.61052632 0.24210526]

[0.10084034 0.21848739 0.68067227]]





precision recall f1-score support

hate 0.13 0.32 0.19 41

offensive 0.25 0.61 0.35 95

neither 0.94 0.68 0.79 714

accuracy 0.66 850

macro avg 0.44 0.54 0.44 850

weighted avg 0.82 0.66 0.71 850

average eval\_loss: 0.8928785784126714, accuracy: 0.6552941176470588

training 8 epoch...

9.34308385848999 seconds used......

8 training finished! average train loss: 0.13200387731194496

total clf loss: 4.7028670236468315 total adv loss: 0.02463627988981898

evaluating...

Confusion matrix, without normalization

[[ 13 17 11]

[ 13 59 23]

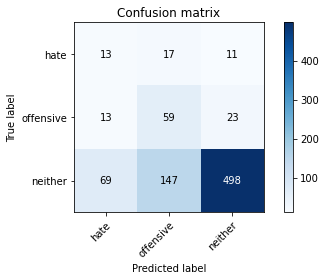
[ 69 147 498]]

Normalized confusion matrix

[[0.31707317 0.41463415 0.26829268]

[0.13684211 0.62105263 0.24210526]

[0.09663866 0.20588235 0.69747899]]





precision recall f1-score support

hate 0.14 0.32 0.19 41

offensive 0.26 0.62 0.37 95

neither 0.94 0.70 0.80 714

accuracy 0.67 850

macro avg 0.45 0.55 0.45 850

weighted avg 0.82 0.67 0.72 850

average eval\_loss: 0.9120018755107425, accuracy: 0.6705882352941176

training 9 epoch...

9.377794742584229 seconds used......

9 training finished! average train loss: 0.09804298123344779

total clf loss: 3.5146894697099924 total adv loss: 0.007428934786730679

evaluating...

Confusion matrix, without normalization

[[ 15 16 10]

[ 15 69 11]

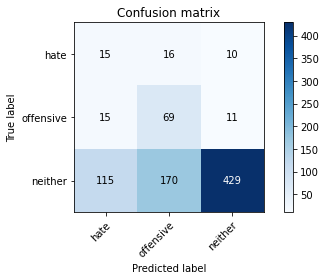
[115 170 429]]

Normalized confusion matrix

[[0.36585366 0.3902439 0.24390244]

[0.15789474 0.72631579 0.11578947]

[0.16106443 0.23809524 0.60084034]]





precision recall f1-score support

hate 0.10 0.37 0.16 41

offensive 0.27 0.73 0.39 95

neither 0.95 0.60 0.74 714

accuracy 0.60 850

macro avg 0.44 0.56 0.43 850

weighted avg 0.84 0.60 0.67 850

average eval\_loss: 1.194365059393783, accuracy: 0.6035294117647059

training 10 epoch...

9.445034265518188 seconds used......

10 training finished! average train loss: 0.06527603403406425

total clf loss: 2.3399807438254356 total adv loss: 0.0049782364053498895

evaluating...

Confusion matrix, without normalization

[[ 9 18 14]

[ 10 67 18]

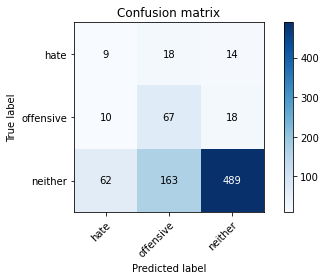
[ 62 163 489]]

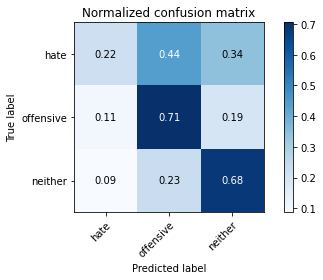
Normalized confusion matrix

[[0.2195122 0.43902439 0.34146341]

[0.10526316 0.70526316 0.18947368]

[0.08683473 0.22829132 0.68487395]]





precision recall f1-score support

hate 0.11 0.22 0.15 41

offensive 0.27 0.71 0.39 95

neither 0.94 0.68 0.79 714

accuracy 0.66 850

macro avg 0.44 0.54 0.44 850

weighted avg 0.82 0.66 0.72 850

average eval\_loss: 1.0585924569560692, accuracy: 0.6647058823529411

4 run......

training 1 epoch...

9.347627639770508 seconds used......

1 training finished! average train loss: 2.4464810623062982

total clf loss: 39.70401358604431 total adv loss: 24.184652268886566

evaluating...

Confusion matrix, without normalization

[[ 0 39 2]

[ 0 90 5]

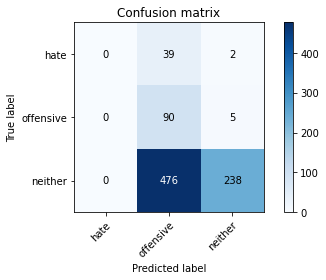
[ 0 476 238]]

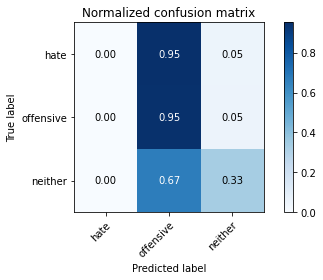
Normalized confusion matrix

[[0. 0.95121951 0.04878049]

[0. 0.94736842 0.05263158]

[0. 0.66666667 0.33333333]]





/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 0.15 0.95 0.26 95

neither 0.97 0.33 0.50 714

accuracy 0.39 850

macro avg 0.37 0.43 0.25 850

weighted avg 0.83 0.39 0.45 850

average eval\_loss: 0.9707102810227594, accuracy: 0.38588235294117645

training 2 epoch...

9.434141159057617 seconds used......

2 training finished! average train loss: 2.0611603458722434

total clf loss: 32.676617830991745 total adv loss: 20.762577384710312

evaluating...

Confusion matrix, without normalization

[[ 0 31 10]

[ 0 66 29]

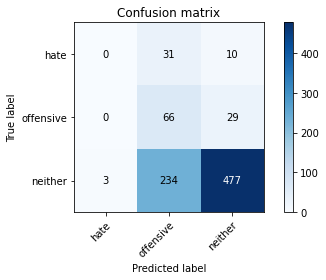
[ 3 234 477]]

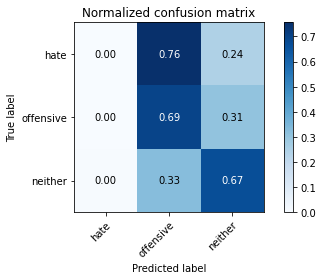
Normalized confusion matrix

[[0. 0.75609756 0.24390244]

[0. 0.69473684 0.30526316]

[0.00420168 0.32773109 0.66806723]]





precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.20 0.69 0.31 95

neither 0.92 0.67 0.78 714

accuracy 0.64 850

macro avg 0.37 0.45 0.36 850

weighted avg 0.80 0.64 0.69 850

average eval\_loss: 0.8425115603347157, accuracy: 0.6388235294117647

training 3 epoch...

9.354011297225952 seconds used......

3 training finished! average train loss: 1.7049610747231378

total clf loss: 27.836263298988342 total adv loss: 16.771167665719986

evaluating...

Confusion matrix, without normalization

[[ 2 23 16]

[ 1 55 39]

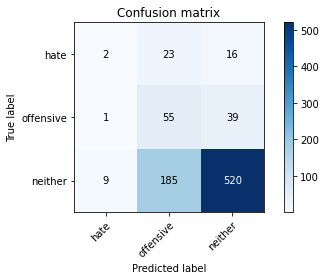
[ 9 185 520]]

Normalized confusion matrix

[[0.04878049 0.56097561 0.3902439 ]

[0.01052632 0.57894737 0.41052632]

[0.01260504 0.25910364 0.72829132]]





precision recall f1-score support

hate 0.17 0.05 0.08 41

offensive 0.21 0.58 0.31 95

neither 0.90 0.73 0.81 714

accuracy 0.68 850

macro avg 0.43 0.45 0.40 850

weighted avg 0.79 0.68 0.72 850

average eval\_loss: 0.800302293931329, accuracy: 0.6788235294117647

training 4 epoch...

9.303058385848999 seconds used......

4 training finished! average train loss: 1.1963554380668535

total clf loss: 21.617777079343796 total adv loss: 10.72550930455327

evaluating...

Confusion matrix, without normalization

[[ 12 20 9]

[ 18 58 19]

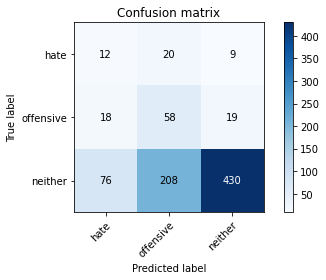
[ 76 208 430]]

Normalized confusion matrix

[[0.29268293 0.48780488 0.2195122 ]

[0.18947368 0.61052632 0.2 ]

[0.10644258 0.29131653 0.6022409 ]]





precision recall f1-score support

hate 0.11 0.29 0.16 41

offensive 0.20 0.61 0.30 95

neither 0.94 0.60 0.73 714

accuracy 0.59 850

macro avg 0.42 0.50 0.40 850

weighted avg 0.82 0.59 0.66 850

average eval\_loss: 0.9406610966421837, accuracy: 0.5882352941176471

training 5 epoch...

9.362797975540161 seconds used......

5 training finished! average train loss: 0.7055386553208033

total clf loss: 15.512060448527336 total adv loss: 4.943665601313114

evaluating...

Confusion matrix, without normalization

[[ 12 17 12]

[ 14 50 31]

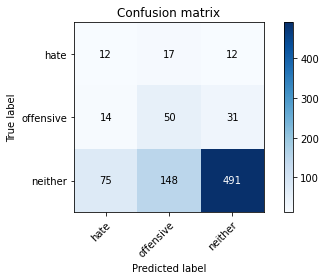
[ 75 148 491]]

Normalized confusion matrix

[[0.29268293 0.41463415 0.29268293]

[0.14736842 0.52631579 0.32631579]

[0.10504202 0.20728291 0.68767507]]





precision recall f1-score support

hate 0.12 0.29 0.17 41

offensive 0.23 0.53 0.32 95

neither 0.92 0.69 0.79 714

accuracy 0.65 850

macro avg 0.42 0.50 0.43 850

weighted avg 0.80 0.65 0.71 850

average eval\_loss: 0.8564464764366316, accuracy: 0.6505882352941177

training 6 epoch...

9.3686363697052 seconds used......

6 training finished! average train loss: 0.38717517794834244

total clf loss: 10.931264877319336 total adv loss: 1.5035207767505199

evaluating...

Confusion matrix, without normalization

[[ 9 22 10]

[ 13 70 12]

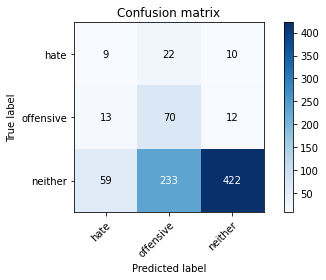
[ 59 233 422]]

Normalized confusion matrix

[[0.2195122 0.53658537 0.24390244]

[0.13684211 0.73684211 0.12631579]

[0.08263305 0.32633053 0.59103641]]





precision recall f1-score support

hate 0.11 0.22 0.15 41

offensive 0.22 0.74 0.33 95

neither 0.95 0.59 0.73 714

accuracy 0.59 850

macro avg 0.43 0.52 0.40 850

weighted avg 0.83 0.59 0.66 850

average eval\_loss: 1.0801766469728116, accuracy: 0.5894117647058823

training 7 epoch...

9.520108699798584 seconds used......

7 training finished! average train loss: 0.22509525985353523

total clf loss: 7.641420356929302 total adv loss: 0.23100449342746288

evaluating...

Confusion matrix, without normalization

[[ 14 17 10]

[ 16 64 15]

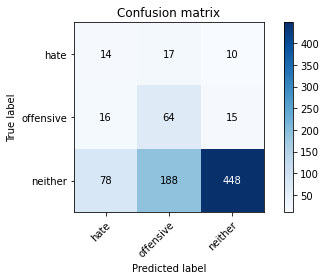
[ 78 188 448]]

Normalized confusion matrix

[[0.34146341 0.41463415 0.24390244]

[0.16842105 0.67368421 0.15789474]

[0.1092437 0.26330532 0.62745098]]





precision recall f1-score support

hate 0.13 0.34 0.19 41

offensive 0.24 0.67 0.35 95

neither 0.95 0.63 0.75 714

accuracy 0.62 850

macro avg 0.44 0.55 0.43 850

weighted avg 0.83 0.62 0.68 850

average eval\_loss: 1.1237764043170353, accuracy: 0.6188235294117647

training 8 epoch...

9.365559339523315 seconds used......

8 training finished! average train loss: 0.16435231185621685

total clf loss: 5.835498556494713 total adv loss: 0.04059232966392301

evaluating...

Confusion matrix, without normalization

[[ 14 18 9]

[ 17 65 13]

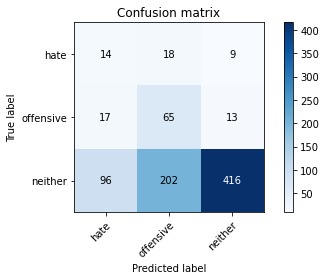
[ 96 202 416]]

Normalized confusion matrix

[[0.34146341 0.43902439 0.2195122 ]

[0.17894737 0.68421053 0.13684211]

[0.13445378 0.28291317 0.58263305]]





precision recall f1-score support

hate 0.11 0.34 0.17 41

offensive 0.23 0.68 0.34 95

neither 0.95 0.58 0.72 714

accuracy 0.58 850

macro avg 0.43 0.54 0.41 850

weighted avg 0.83 0.58 0.65 850

average eval\_loss: 1.273907339936772, accuracy: 0.5823529411764706

training 9 epoch...

9.361944675445557 seconds used......

9 training finished! average train loss: 0.12304563416788976

total clf loss: 4.404929142910987 total adv loss: 0.01235684280982241

evaluating...

Confusion matrix, without normalization

[[ 14 17 10]

[ 19 62 14]

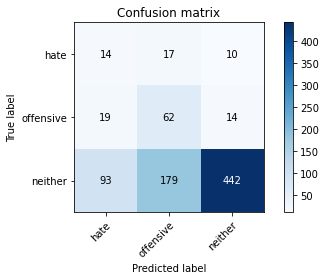
[ 93 179 442]]

Normalized confusion matrix

[[0.34146341 0.41463415 0.24390244]

[0.2 0.65263158 0.14736842]

[0.1302521 0.25070028 0.61904762]]





precision recall f1-score support

hate 0.11 0.34 0.17 41

offensive 0.24 0.65 0.35 95

neither 0.95 0.62 0.75 714

accuracy 0.61 850

macro avg 0.43 0.54 0.42 850

weighted avg 0.83 0.61 0.68 850

average eval\_loss: 1.2669079687532991, accuracy: 0.6094117647058823

training 10 epoch...

9.353253841400146 seconds used......

10 training finished! average train loss: 0.08442935306164953

total clf loss: 3.0066487435251474 total adv loss: 0.016403983130658162

evaluating...

Confusion matrix, without normalization

[[ 13 16 12]

[ 12 54 29]

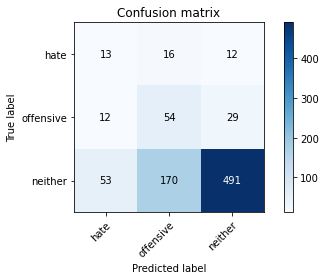
[ 53 170 491]]

Normalized confusion matrix

[[0.31707317 0.3902439 0.29268293]

[0.12631579 0.56842105 0.30526316]

[0.07422969 0.23809524 0.68767507]]





precision recall f1-score support

hate 0.17 0.32 0.22 41

offensive 0.23 0.57 0.32 95

neither 0.92 0.69 0.79 714

accuracy 0.66 850

macro avg 0.44 0.52 0.44 850

weighted avg 0.81 0.66 0.71 850

average eval\_loss: 1.2179131503216922, accuracy: 0.6564705882352941

5 run......

training 1 epoch...

9.339124917984009 seconds used......

1 training finished! average train loss: 2.4471011691623263

total clf loss: 40.76767712831497 total adv loss: 23.663982182741165

evaluating...

Confusion matrix, without normalization

[[ 0 27 14]

[ 0 32 63]

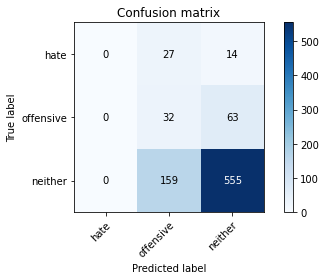
[ 0 159 555]]

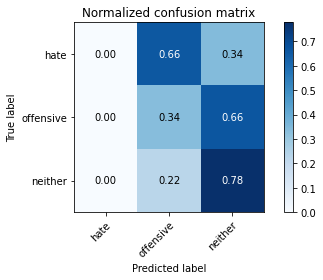
Normalized confusion matrix

[[0. 0.65853659 0.34146341]

[0. 0.33684211 0.66315789]

[0. 0.22268908 0.77731092]]





/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 0.15 0.34 0.20 95

neither 0.88 0.78 0.82 714

accuracy 0.69 850

macro avg 0.34 0.37 0.34 850

weighted avg 0.75 0.69 0.72 850

average eval\_loss: 0.860953400994456, accuracy: 0.6905882352941176

training 2 epoch...

9.40485429763794 seconds used......

2 training finished! average train loss: 2.113966792821884

total clf loss: 34.54585462808609 total adv loss: 20.77847495675087

evaluating...

Confusion matrix, without normalization

[[ 0 28 13]

[ 0 57 38]

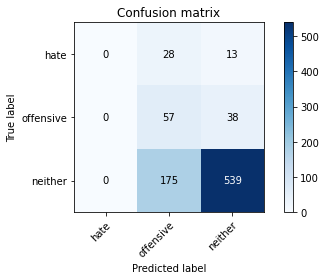
[ 0 175 539]]

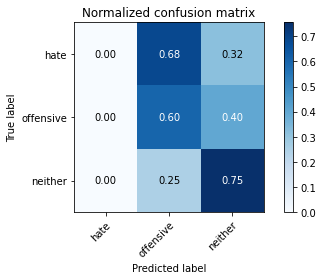
Normalized confusion matrix

[[0. 0.68292683 0.31707317]

[0. 0.6 0.4 ]

[0. 0.24509804 0.75490196]]





/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 0.22 0.60 0.32 95

neither 0.91 0.75 0.83 714

accuracy 0.70 850

macro avg 0.38 0.45 0.38 850

weighted avg 0.79 0.70 0.73 850

average eval\_loss: 0.7798337180947148, accuracy: 0.7011764705882353

training 3 epoch...

9.352848291397095 seconds used......

3 training finished! average train loss: 1.8374791509575314

total clf loss: 30.373204469680786 total adv loss: 17.88802269101143

evaluating...

Confusion matrix, without normalization

[[ 0 25 16]

[ 0 59 36]

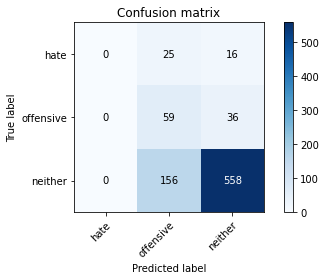
[ 0 156 558]]

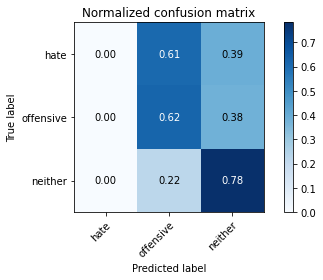
Normalized confusion matrix

[[0. 0.6097561 0.3902439 ]

[0. 0.62105263 0.37894737]

[0. 0.21848739 0.78151261]]





/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 0.25 0.62 0.35 95

neither 0.91 0.78 0.84 714

accuracy 0.73 850

macro avg 0.39 0.47 0.40 850

weighted avg 0.80 0.73 0.75 850

average eval\_loss: 0.7691116488950197, accuracy: 0.7258823529411764

training 4 epoch...

9.488605976104736 seconds used......

4 training finished! average train loss: 1.3873865455389023

total clf loss: 25.1585873067379 total adv loss: 12.393664218485355

evaluating...

Confusion matrix, without normalization

[[ 2 18 21]

[ 0 56 39]

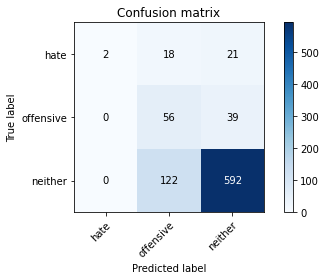
[ 0 122 592]]

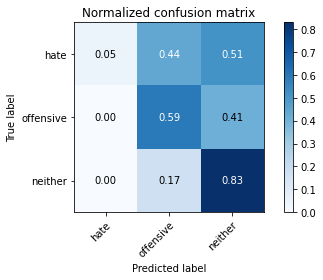
Normalized confusion matrix

[[0.04878049 0.43902439 0.51219512]

[0. 0.58947368 0.41052632]

[0. 0.17086835 0.82913165]]





precision recall f1-score support

hate 1.00 0.05 0.09 41

offensive 0.29 0.59 0.38 95

neither 0.91 0.83 0.87 714

accuracy 0.76 850

macro avg 0.73 0.49 0.45 850

weighted avg 0.84 0.76 0.78 850

average eval\_loss: 0.6746814431146134, accuracy: 0.7647058823529411

training 5 epoch...

9.331871032714844 seconds used......

5 training finished! average train loss: 0.9892224280370606

total clf loss: 20.239775329828262 total adv loss: 7.6861160565167665

evaluating...

Confusion matrix, without normalization

[[ 5 17 19]

[ 3 66 26]

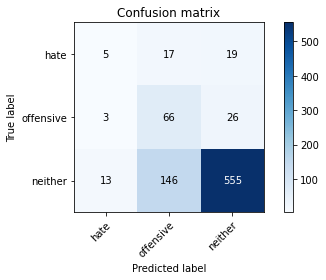
[ 13 146 555]]

Normalized confusion matrix

[[0.12195122 0.41463415 0.46341463]

[0.03157895 0.69473684 0.27368421]

[0.01820728 0.20448179 0.77731092]]





precision recall f1-score support

hate 0.24 0.12 0.16 41

offensive 0.29 0.69 0.41 95

neither 0.93 0.78 0.84 714

accuracy 0.74 850

macro avg 0.48 0.53 0.47 850

weighted avg 0.82 0.74 0.76 850

average eval\_loss: 0.714010454021221, accuracy: 0.7364705882352941

training 6 epoch...

9.480894327163696 seconds used......

6 training finished! average train loss: 0.587669012033277

total clf loss: 14.694696769118309 total adv loss: 3.230693868594244

evaluating...

Confusion matrix, without normalization

[[ 8 16 17]

[ 10 67 18]

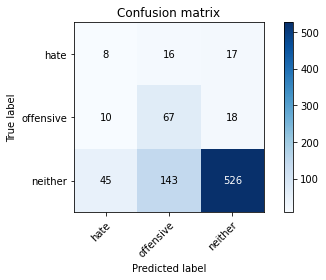
[ 45 143 526]]

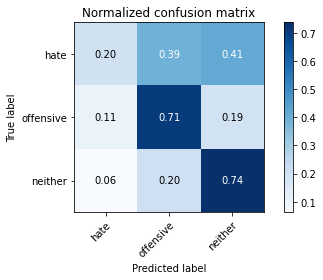
Normalized confusion matrix

[[0.19512195 0.3902439 0.41463415]

[0.10526316 0.70526316 0.18947368]

[0.06302521 0.20028011 0.73669468]]





precision recall f1-score support

hate 0.13 0.20 0.15 41

offensive 0.30 0.71 0.42 95

neither 0.94 0.74 0.83 714

accuracy 0.71 850

macro avg 0.45 0.55 0.47 850

weighted avg 0.83 0.71 0.75 850

average eval\_loss: 0.7580398163476656, accuracy: 0.7070588235294117

training 7 epoch...

9.360025405883789 seconds used......

7 training finished! average train loss: 0.33177927053636974

total clf loss: 10.248461671173573 total adv loss: 0.8477960273157805

evaluating...

Confusion matrix, without normalization

[[ 10 20 11]

[ 13 72 10]

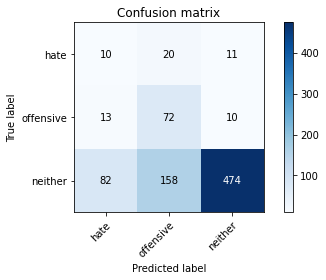
[ 82 158 474]]

Normalized confusion matrix

[[0.24390244 0.48780488 0.26829268]

[0.13684211 0.75789474 0.10526316]

[0.11484594 0.22128852 0.66386555]]





precision recall f1-score support

hate 0.10 0.24 0.14 41

offensive 0.29 0.76 0.42 95

neither 0.96 0.66 0.78 714

accuracy 0.65 850

macro avg 0.45 0.56 0.45 850

weighted avg 0.84 0.65 0.71 850

average eval\_loss: 0.8914530371683974, accuracy: 0.6541176470588236

training 8 epoch...

9.358933210372925 seconds used......

8 training finished! average train loss: 0.24997237148798174

total clf loss: 7.868299428373575 total adv loss: 0.5653529803239508

evaluating...

Confusion matrix, without normalization

[[ 11 18 12]

[ 13 64 18]

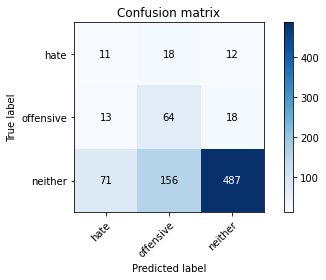
[ 71 156 487]]

Normalized confusion matrix

[[0.26829268 0.43902439 0.29268293]

[0.13684211 0.67368421 0.18947368]

[0.09943978 0.21848739 0.68207283]]





precision recall f1-score support

hate 0.12 0.27 0.16 41

offensive 0.27 0.67 0.38 95

neither 0.94 0.68 0.79 714

accuracy 0.66 850

macro avg 0.44 0.54 0.45 850

weighted avg 0.83 0.66 0.72 850

average eval\_loss: 0.9072947561914144, accuracy: 0.6611764705882353

training 9 epoch...

9.516562223434448 seconds used......

9 training finished! average train loss: 0.19765681131846374

total clf loss: 6.726636663079262 total adv loss: 0.19450425637478475

evaluating...

Confusion matrix, without normalization

[[ 11 18 12]

[ 12 67 16]

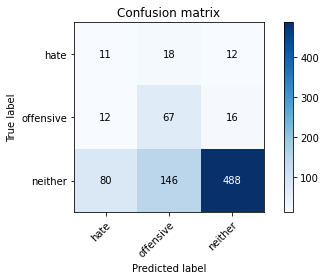
[ 80 146 488]]

Normalized confusion matrix

[[0.26829268 0.43902439 0.29268293]

[0.12631579 0.70526316 0.16842105]

[0.11204482 0.20448179 0.68347339]]





precision recall f1-score support

hate 0.11 0.27 0.15 41

offensive 0.29 0.71 0.41 95

neither 0.95 0.68 0.79 714

accuracy 0.67 850

macro avg 0.45 0.55 0.45 850

weighted avg 0.83 0.67 0.72 850

average eval\_loss: 0.9046604188029156, accuracy: 0.6658823529411765

training 10 epoch...

9.378364086151123 seconds used......

10 training finished! average train loss: 0.12547007327278456

total clf loss: 4.4490615632385015 total adv loss: 0.03393054268963169

evaluating...

Confusion matrix, without normalization

[[ 10 17 14]

[ 11 67 17]

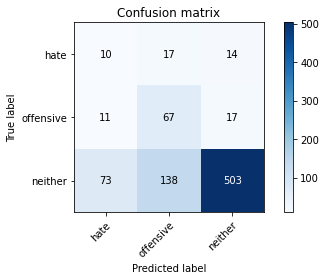
[ 73 138 503]]

Normalized confusion matrix

[[0.24390244 0.41463415 0.34146341]

[0.11578947 0.70526316 0.17894737]

[0.1022409 0.19327731 0.70448179]]





precision recall f1-score support

hate 0.11 0.24 0.15 41

offensive 0.30 0.71 0.42 95

neither 0.94 0.70 0.81 714

accuracy 0.68 850

macro avg 0.45 0.55 0.46 850

weighted avg 0.83 0.68 0.73 850

average eval\_loss: 0.9120069106737542, accuracy: 0.6823529411764706