hate as no-offensive

alpha=2

training 1 epoch...

85.08301401138306 seconds used......

1 training finished! average train loss: 0.9322657774007598

total clf loss: 68.96710445731878 total adv loss: 45.69130452349782

evaluating...

Confusion matrix, without normalization

[[ 0 0 13]

[ 0 0 29]

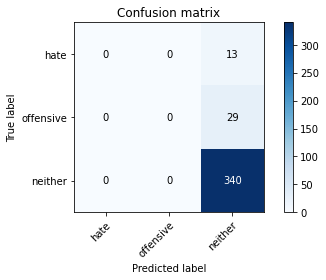
[ 0 0 340]]

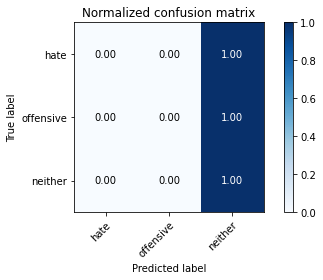
Normalized confusion matrix

[[0. 0. 1.]

[0. 0. 1.]

[0. 0. 1.]]





/usr/local/lib/python3.7/dist-packages/sklearn/metrics/\_classification.py:1272: 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 13

offensive 0.00 0.00 0.00 29

neither 0.89 1.00 0.94 340

accuracy 0.89 382

macro avg 0.30 0.33 0.31 382

weighted avg 0.79 0.89 0.84 382

average eval\_loss: 0.4128064729273319, accuracy: 0.8900523560209425

training 2 epoch...

84.22950768470764 seconds used......

2 training finished! average train loss: 0.7263237367882285

total clf loss: 53.493263605982065 total adv loss: 35.71720964182168

evaluating...

Confusion matrix, without normalization

[[ 0 0 13]

[ 0 0 29]

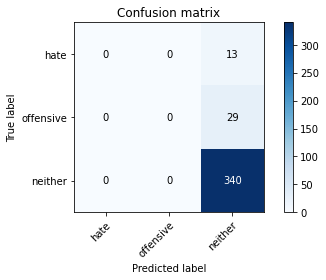
[ 0 0 340]]

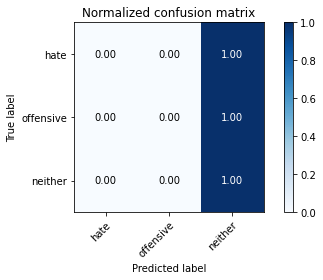
Normalized confusion matrix

[[0. 0. 1.]

[0. 0. 1.]

[0. 0. 1.]]





precision recall f1-score support

hate 0.00 0.00 0.00 13

offensive 0.00 0.00 0.00 29

neither 0.89 1.00 0.94 340

accuracy 0.89 382

macro avg 0.30 0.33 0.31 382

weighted avg 0.79 0.89 0.84 382

average eval\_loss: 0.5179981334134937, accuracy: 0.8900523560209425

training 3 epoch...

84.0239155292511 seconds used......

3 training finished! average train loss: 0.44011329820509565

total clf loss: 35.11405636742711 total adv loss: 20.292715487477835

evaluating...

Confusion matrix, without normalization

[[ 0 3 10]

[ 0 3 26]

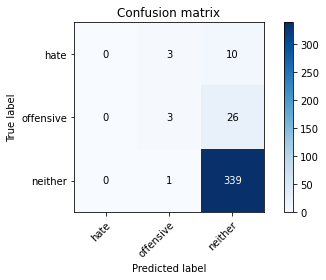
[ 0 1 339]]

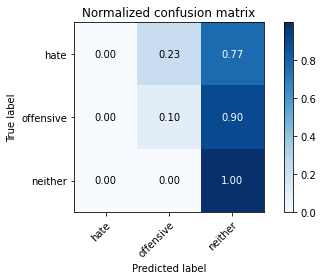
Normalized confusion matrix

[[0. 0.23076923 0.76923077]

[0. 0.10344828 0.89655172]

[0. 0.00294118 0.99705882]]





precision recall f1-score support

hate 0.00 0.00 0.00 13

offensive 0.43 0.10 0.17 29

neither 0.90 1.00 0.95 340

accuracy 0.90 382

macro avg 0.44 0.37 0.37 382

weighted avg 0.84 0.90 0.86 382

average eval\_loss: 0.3530836520716548, accuracy: 0.8952879581151832

training 4 epoch...

84.09191012382507 seconds used......

4 training finished! average train loss: 0.19970864984531736

total clf loss: 20.20256929565221 total adv loss: 7.07365935305279

evaluating...

Confusion matrix, without normalization

[[ 0 2 11]

[ 0 5 24]

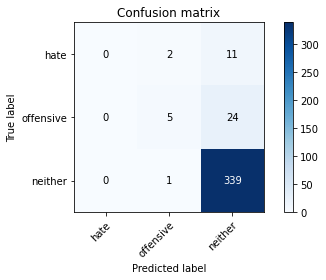
[ 0 1 339]]

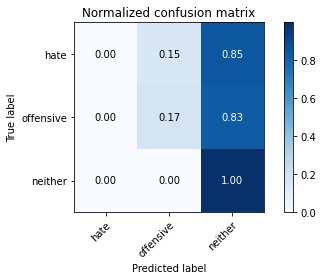
Normalized confusion matrix

[[0. 0.15384615 0.84615385]

[0. 0.17241379 0.82758621]

[0. 0.00294118 0.99705882]]





precision recall f1-score support

hate 0.00 0.00 0.00 13

offensive 0.62 0.17 0.27 29

neither 0.91 1.00 0.95 340

accuracy 0.90 382

macro avg 0.51 0.39 0.41 382

weighted avg 0.85 0.90 0.87 382

average eval\_loss: 0.4452547195716761, accuracy: 0.900523560209424

training 5 epoch...

83.94765448570251 seconds used......

5 training finished! average train loss: 0.09454772257256906

total clf loss: 12.990880495868623 total adv loss: 1.6356638636351875

evaluating...

Confusion matrix, without normalization

[[ 2 2 9]

[ 0 5 24]

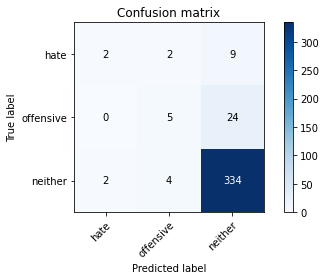
[ 2 4 334]]

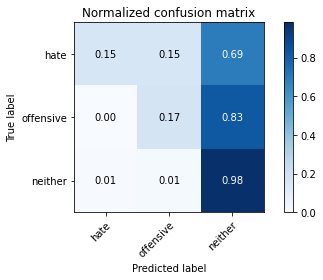
Normalized confusion matrix

[[0.15384615 0.15384615 0.69230769]

[0. 0.17241379 0.82758621]

[0.00588235 0.01176471 0.98235294]]





precision recall f1-score support

hate 0.50 0.15 0.24 13

offensive 0.45 0.17 0.25 29

neither 0.91 0.98 0.94 340

accuracy 0.89 382

macro avg 0.62 0.44 0.48 382

weighted avg 0.86 0.89 0.87 382

average eval\_loss: 0.5555694501847028, accuracy: 0.8926701570680629

model of 5 epoch

Confusion matrix, without normalization

[[ 0 1 40]

[ 4 4 87]

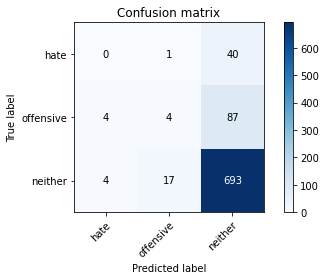
[ 4 17 693]]

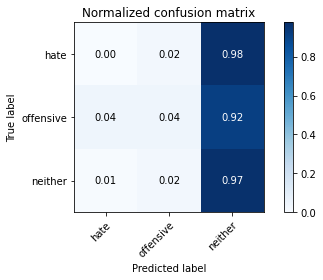
Normalized confusion matrix

[[0. 0.02439024 0.97560976]

[0.04210526 0.04210526 0.91578947]

[0.00560224 0.02380952 0.97058824]]





precision recall f1-score support

hate 0.00 0.00 0.00 41

offensive 0.18 0.04 0.07 95

neither 0.85 0.97 0.90 714

accuracy 0.82 850

macro avg 0.34 0.34 0.32 850

weighted avg 0.73 0.82 0.77 850

average eval\_loss: 0.8789495398983532, accuracy: 0.82

0.82