hate as no-offensive

alpha=2

training 1 epoch...

115.22534275054932 seconds used......

1 training finished! average train loss: 1.7583423567633343

total clf loss: 170.97959205508232 total adv loss: 120.23625968396664

evaluating...

Confusion matrix, without normalization

[[ 0 20 25]

[ 0 121 65]

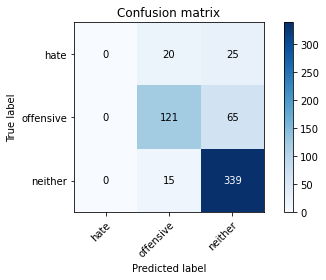
[ 0 15 339]]

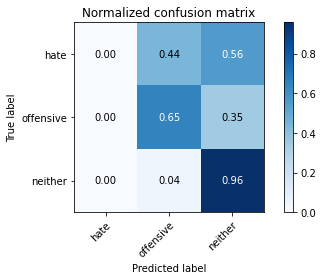
Normalized confusion matrix

[[0. 0.44444444 0.55555556]

[0. 0.65053763 0.34946237]

[0. 0.04237288 0.95762712]]





/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 45

offensive 0.78 0.65 0.71 186

neither 0.79 0.96 0.87 354

accuracy 0.79 585

macro avg 0.52 0.54 0.52 585

weighted avg 0.72 0.79 0.75 585

average eval\_loss: 0.5504734059174855, accuracy: 0.7863247863247863

training 2 epoch...

114.45849800109863 seconds used......

2 training finished! average train loss: 1.2660781304296265

total clf loss: 122.45309748500586 total adv loss: 86.90459240972996

evaluating...

Confusion matrix, without normalization

[[ 1 22 22]

[ 0 136 50]

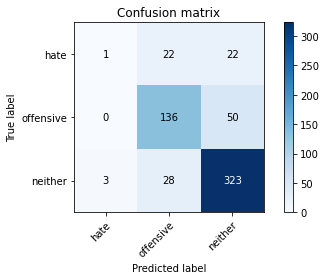
[ 3 28 323]]

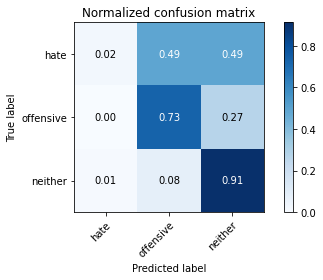
Normalized confusion matrix

[[0.02222222 0.48888889 0.48888889]

[0. 0.7311828 0.2688172 ]

[0.00847458 0.07909605 0.91242938]]





precision recall f1-score support

hate 0.25 0.02 0.04 45

offensive 0.73 0.73 0.73 186

neither 0.82 0.91 0.86 354

accuracy 0.79 585

macro avg 0.60 0.56 0.54 585

weighted avg 0.75 0.79 0.76 585

average eval\_loss: 0.5164935737848282, accuracy: 0.7863247863247863

training 3 epoch...

114.30922985076904 seconds used......

3 training finished! average train loss: 0.8651482462882996

total clf loss: 84.97103249281645 total adv loss: 58.73682865686715

evaluating...

Confusion matrix, without normalization

[[ 6 23 16]

[ 3 152 31]

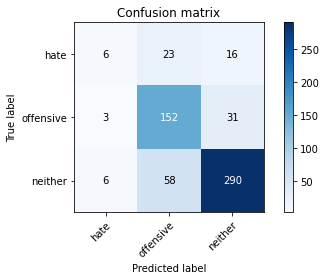
[ 6 58 290]]

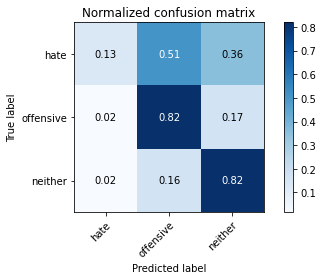
Normalized confusion matrix

[[0.13333333 0.51111111 0.35555556]

[0.01612903 0.8172043 0.16666667]

[0.01694915 0.16384181 0.81920904]]





precision recall f1-score support

hate 0.40 0.13 0.20 45

offensive 0.65 0.82 0.73 186

neither 0.86 0.82 0.84 354

accuracy 0.77 585

macro avg 0.64 0.59 0.59 585

weighted avg 0.76 0.77 0.75 585

average eval\_loss: 0.548185637096564, accuracy: 0.7658119658119659

training 4 epoch...

114.28819727897644 seconds used......

4 training finished! average train loss: 0.539938217586177

total clf loss: 55.98347897455096 total adv loss: 35.18103201221675

evaluating...

Confusion matrix, without normalization

[[ 7 17 21]

[ 2 130 54]

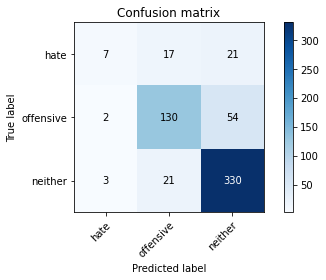
[ 3 21 330]]

Normalized confusion matrix

[[0.15555556 0.37777778 0.46666667]

[0.01075269 0.69892473 0.29032258]

[0.00847458 0.05932203 0.93220339]]





precision recall f1-score support

hate 0.58 0.16 0.25 45

offensive 0.77 0.70 0.73 186

neither 0.81 0.93 0.87 354

accuracy 0.80 585

macro avg 0.72 0.60 0.62 585

weighted avg 0.78 0.80 0.78 585

average eval\_loss: 0.6169950279096762, accuracy: 0.7982905982905983

training 5 epoch...

114.37676119804382 seconds used......

5 training finished! average train loss: 0.3554824706342103

total clf loss: 37.663081613369286 total adv loss: 22.759908285923302

evaluating...

Confusion matrix, without normalization

[[ 15 13 17]

[ 17 117 52]

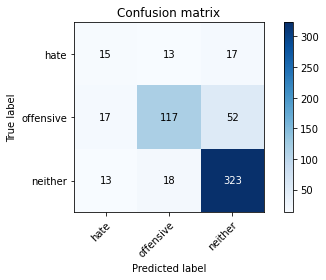
[ 13 18 323]]

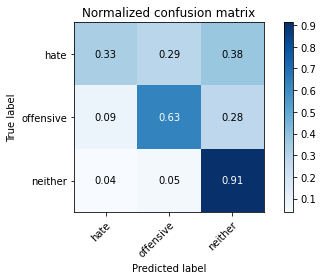
Normalized confusion matrix

[[0.33333333 0.28888889 0.37777778]

[0.09139785 0.62903226 0.27956989]

[0.03672316 0.05084746 0.91242938]]





precision recall f1-score support

hate 0.33 0.33 0.33 45

offensive 0.79 0.63 0.70 186

neither 0.82 0.91 0.87 354

accuracy 0.78 585

macro avg 0.65 0.62 0.63 585

weighted avg 0.78 0.78 0.77 585

average eval\_loss: 0.6910962065060934, accuracy: 0.7777777777777778

model of 5 epoch

Confusion matrix, without normalization

[[ 26 11 14]

[ 18 103 48]

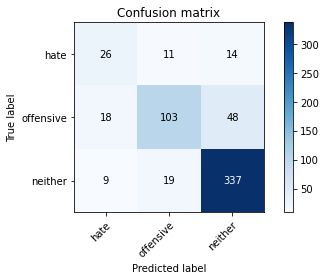
[ 9 19 337]]

Normalized confusion matrix

[[0.50980392 0.21568627 0.2745098 ]

[0.10650888 0.60946746 0.28402367]

[0.02465753 0.05205479 0.92328767]]





precision recall f1-score support

hate 0.49 0.51 0.50 51

offensive 0.77 0.61 0.68 169

neither 0.84 0.92 0.88 365

accuracy 0.80 585

macro avg 0.70 0.68 0.69 585

weighted avg 0.79 0.80 0.79 585

average eval\_loss: 0.7179318773249785, accuracy: 0.7965811965811965

0.7965811965811965