hate as no-offensive

alpha=2

training 1 epoch...

406.24265122413635 seconds used......

1 training finished! average train loss: 1.674773651993636

total clf loss: 225.0465122461319 total adv loss: 108.54686599224806

evaluating...

Confusion matrix, without normalization

[[ 0 9 112]

[ 0 47 60]

[ 0 27 331]]

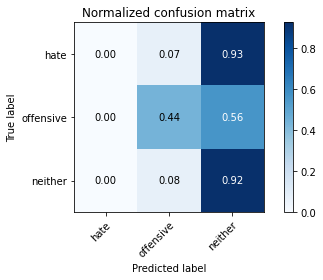
Normalized confusion matrix

[[0. 0.07438017 0.92561983]

[0. 0.43925234 0.56074766]

[0. 0.07541899 0.92458101]]





/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 0.00 121

offensive 0.57 0.44 0.49 107

neither 0.66 0.92 0.77 358

accuracy 0.65 586

macro avg 0.41 0.45 0.42 586

weighted avg 0.51 0.65 0.56 586

average eval\_loss: 0.8017523348331451, accuracy: 0.6450511945392492

training 2 epoch...

404.95077657699585 seconds used......

2 training finished! average train loss: 1.3854566144220757

total clf loss: 194.07431051135063 total adv loss: 85.84311812743545

evaluating...

Confusion matrix, without normalization

[[ 0 7 114]

[ 0 39 68]

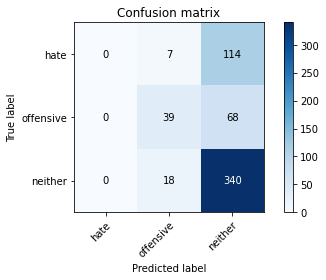
[ 0 18 340]]

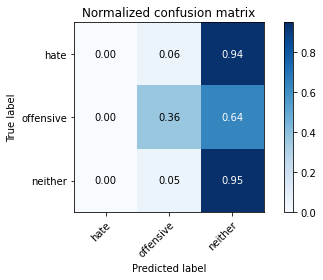
Normalized confusion matrix

[[0. 0.05785124 0.94214876]

[0. 0.36448598 0.63551402]

[0. 0.05027933 0.94972067]]





precision recall f1-score support

hate 0.00 0.00 0.00 121

offensive 0.61 0.36 0.46 107

neither 0.65 0.95 0.77 358

accuracy 0.65 586

macro avg 0.42 0.44 0.41 586

weighted avg 0.51 0.65 0.56 586

average eval\_loss: 0.800417164961497, accuracy: 0.6467576791808873

training 3 epoch...

404.9791543483734 seconds used......

3 training finished! average train loss: 1.0893428671089085

total clf loss: 164.52018281817436 total adv loss: 61.53316760621965

evaluating...

Confusion matrix, without normalization

[[ 20 13 88]

[ 8 45 54]

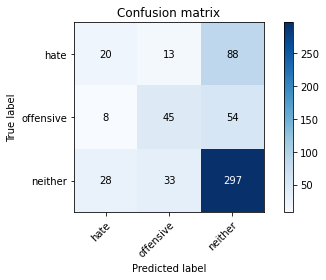
[ 28 33 297]]

Normalized confusion matrix

[[0.16528926 0.10743802 0.72727273]

[0.07476636 0.42056075 0.5046729 ]

[0.07821229 0.09217877 0.82960894]]





precision recall f1-score support

hate 0.36 0.17 0.23 121

offensive 0.49 0.42 0.45 107

neither 0.68 0.83 0.75 358

accuracy 0.62 586

macro avg 0.51 0.47 0.48 586

weighted avg 0.58 0.62 0.58 586

average eval\_loss: 0.8587170124053956, accuracy: 0.6177474402730375

training 4 epoch...

405.1707134246826 seconds used......

4 training finished! average train loss: 0.712096014878515

total clf loss: 125.21556852757931 total adv loss: 31.388889664551243

evaluating...

Confusion matrix, without normalization

[[ 52 19 50]

[ 12 57 38]

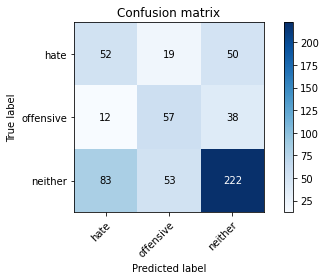
[ 83 53 222]]

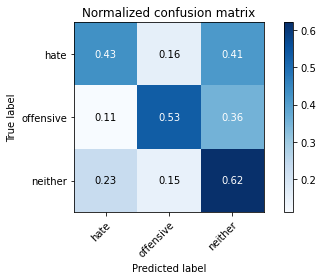
Normalized confusion matrix

[[0.42975207 0.15702479 0.41322314]

[0.11214953 0.53271028 0.35514019]

[0.23184358 0.14804469 0.62011173]]





precision recall f1-score support

hate 0.35 0.43 0.39 121

offensive 0.44 0.53 0.48 107

neither 0.72 0.62 0.66 358

accuracy 0.56 586

macro avg 0.50 0.53 0.51 586

weighted avg 0.59 0.56 0.57 586

average eval\_loss: 1.0389453500509263, accuracy: 0.5648464163822525

training 5 epoch...

404.2787778377533 seconds used......

5 training finished! average train loss: 0.49215095301130507

total clf loss: 94.93397611379623 total adv loss: 17.4969377637899

evaluating...

Confusion matrix, without normalization

[[ 37 14 70]

[ 19 48 40]

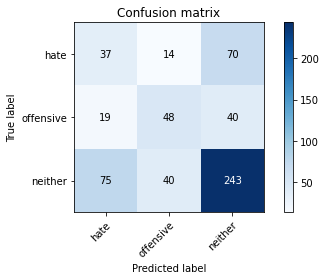
[ 75 40 243]]

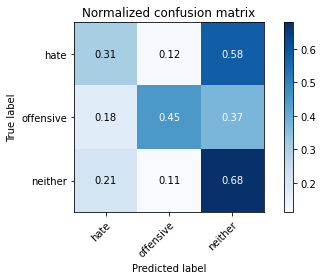
Normalized confusion matrix

[[0.30578512 0.11570248 0.5785124 ]

[0.17757009 0.44859813 0.37383178]

[0.20949721 0.11173184 0.67877095]]





precision recall f1-score support

hate 0.28 0.31 0.29 121

offensive 0.47 0.45 0.46 107

neither 0.69 0.68 0.68 358

accuracy 0.56 586

macro avg 0.48 0.48 0.48 586

weighted avg 0.56 0.56 0.56 586

average eval\_loss: 1.2007285952568054, accuracy: 0.5597269624573379

model of 4 epoch

Confusion matrix, without normalization

[[ 89 16 19]

[ 36 99 29]

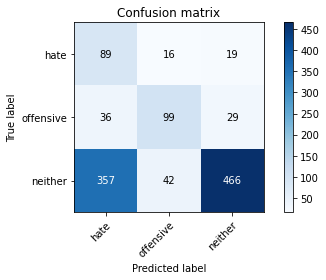
[357 42 466]]

Normalized confusion matrix

[[0.71774194 0.12903226 0.15322581]

[0.2195122 0.60365854 0.17682927]

[0.41271676 0.04855491 0.53872832]]





precision recall f1-score support

hate 0.18 0.72 0.29 124

offensive 0.63 0.60 0.62 164

neither 0.91 0.54 0.68 865

accuracy 0.57 1153

macro avg 0.57 0.62 0.53 1153

weighted avg 0.79 0.57 0.63 1153

average eval\_loss: 0.9213029012597841, accuracy: 0.5672159583694709