hate as no-offensive

alpha=2

training 1 epoch...

120.02953720092773 seconds used......

1 training finished! average train loss: 1.5976926078144302

total clf loss: 152.70516416430473 total adv loss: 110.57745370268822

evaluating...

Confusion matrix, without normalization

[[ 4 22 19]

[ 0 128 58]

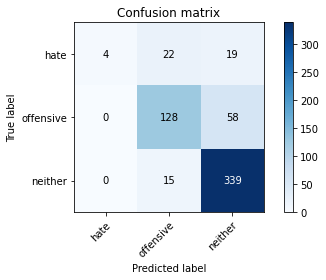
[ 0 15 339]]

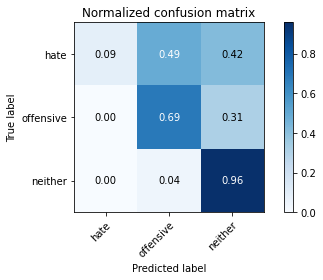
Normalized confusion matrix

[[0.08888889 0.48888889 0.42222222]

[0. 0.68817204 0.31182796]

[0. 0.04237288 0.95762712]]





precision recall f1-score support

hate 1.00 0.09 0.16 45

offensive 0.78 0.69 0.73 186

neither 0.81 0.96 0.88 354

accuracy 0.81 585

macro avg 0.86 0.58 0.59 585

weighted avg 0.82 0.81 0.78 585

average eval\_loss: 0.47327499787012733, accuracy: 0.8051282051282052

training 2 epoch...

119.42105197906494 seconds used......

2 training finished! average train loss: 0.94691216512623

total clf loss: 90.35537360236049 total adv loss: 65.61103632673621

evaluating...

Confusion matrix, without normalization

[[ 7 27 11]

[ 0 157 29]

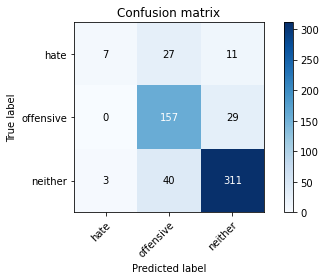
[ 3 40 311]]

Normalized confusion matrix

[[0.15555556 0.6 0.24444444]

[0. 0.84408602 0.15591398]

[0.00847458 0.11299435 0.87853107]]





precision recall f1-score support

hate 0.70 0.16 0.25 45

offensive 0.70 0.84 0.77 186

neither 0.89 0.88 0.88 354

accuracy 0.81 585

macro avg 0.76 0.63 0.63 585

weighted avg 0.81 0.81 0.80 585

average eval\_loss: 0.46560748741030694, accuracy: 0.811965811965812

training 3 epoch...

119.39003252983093 seconds used......

3 training finished! average train loss: 0.4843059429723738

total clf loss: 50.14819122478366 total adv loss: 31.589699518866837

evaluating...

Confusion matrix, without normalization

[[ 25 9 11]

[ 17 133 36]

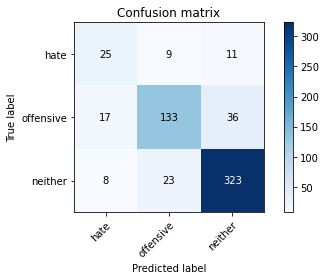
[ 8 23 323]]

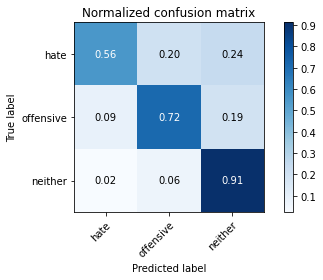
Normalized confusion matrix

[[0.55555556 0.2 0.24444444]

[0.09139785 0.71505376 0.19354839]

[0.02259887 0.06497175 0.91242938]]





precision recall f1-score support

hate 0.50 0.56 0.53 45

offensive 0.81 0.72 0.76 186

neither 0.87 0.91 0.89 354

accuracy 0.82 585

macro avg 0.73 0.73 0.73 585

weighted avg 0.82 0.82 0.82 585

average eval\_loss: 0.5412963186701139, accuracy: 0.8222222222222222

training 4 epoch...

119.34546208381653 seconds used......

4 training finished! average train loss: 0.26065541846024937

total clf loss: 27.168053126893938 total adv loss: 16.912657475157175

evaluating...

Confusion matrix, without normalization

[[ 20 17 8]

[ 9 143 34]

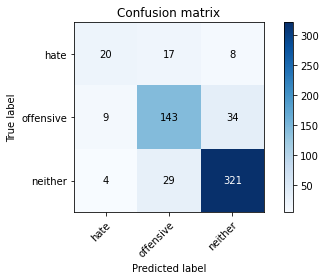
[ 4 29 321]]

Normalized confusion matrix

[[0.44444444 0.37777778 0.17777778]

[0.0483871 0.7688172 0.1827957 ]

[0.01129944 0.0819209 0.90677966]]





precision recall f1-score support

hate 0.61 0.44 0.51 45

offensive 0.76 0.77 0.76 186

neither 0.88 0.91 0.90 354

accuracy 0.83 585

macro avg 0.75 0.71 0.72 585

weighted avg 0.82 0.83 0.82 585

average eval\_loss: 0.6011831281085809, accuracy: 0.8273504273504273

training 5 epoch...

119.49027228355408 seconds used......

5 training finished! average train loss: 0.15842492002635622

total clf loss: 16.437471957178786 total adv loss: 10.31697968498338

evaluating...

Confusion matrix, without normalization

[[ 23 15 7]

[ 21 140 25]

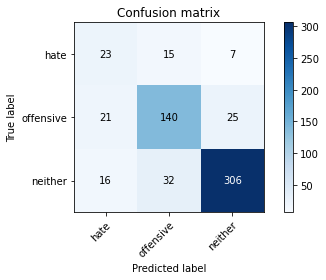
[ 16 32 306]]

Normalized confusion matrix

[[0.51111111 0.33333333 0.15555556]

[0.11290323 0.75268817 0.1344086 ]

[0.04519774 0.09039548 0.86440678]]





precision recall f1-score support

hate 0.38 0.51 0.44 45

offensive 0.75 0.75 0.75 186

neither 0.91 0.86 0.88 354

accuracy 0.80 585

macro avg 0.68 0.71 0.69 585

weighted avg 0.82 0.80 0.81 585

average eval\_loss: 0.698900556564331, accuracy: 0.8017094017094017

model of 3 epoch

Confusion matrix, without normalization

[[ 33 11 7]

[ 17 119 33]

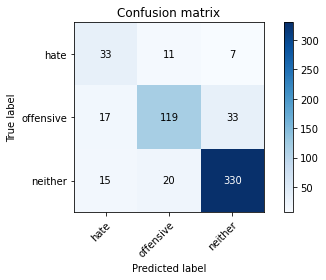
[ 15 20 330]]

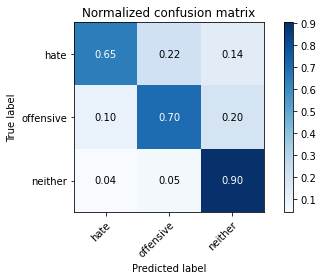
Normalized confusion matrix

[[0.64705882 0.21568627 0.1372549 ]

[0.10059172 0.70414201 0.19526627]

[0.04109589 0.05479452 0.90410959]]





precision recall f1-score support

hate 0.51 0.65 0.57 51

offensive 0.79 0.70 0.75 169

neither 0.89 0.90 0.90 365

accuracy 0.82 585

macro avg 0.73 0.75 0.74 585

weighted avg 0.83 0.82 0.83 585

average eval\_loss: 0.5483509545524915, accuracy: 0.8239316239316239

0.8239316239316239