hate as no-offensive

alpha=2

training 1 epoch...

172.84667015075684 seconds used......

1 training finished! average train loss: 1.4927446918828147

total clf loss: 140.06251746416092 total adv loss: 86.70693423599005

evaluating...

Confusion matrix, without normalization

[[ 3 20 30]

[ 5 147 29]

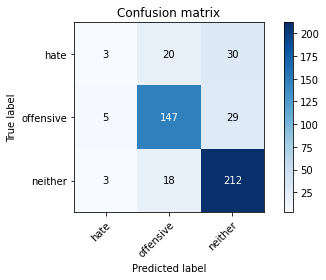
[ 3 18 212]]

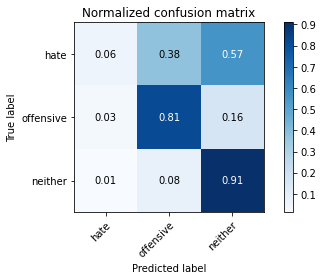
Normalized confusion matrix

[[0.05660377 0.37735849 0.56603774]

[0.02762431 0.8121547 0.16022099]

[0.01287554 0.07725322 0.90987124]]





precision recall f1-score support

hate 0.27 0.06 0.09 53

offensive 0.79 0.81 0.80 181

neither 0.78 0.91 0.84 233

accuracy 0.78 467

macro avg 0.62 0.59 0.58 467

weighted avg 0.73 0.78 0.74 467

average eval\_loss: 0.5382552991310755, accuracy: 0.7751605995717344

training 2 epoch...

172.25479412078857 seconds used......

2 training finished! average train loss: 1.0370415376765387

total clf loss: 98.72259613871574 total adv loss: 59.528063371777534

evaluating...

Confusion matrix, without normalization

[[ 11 17 25]

[ 16 145 20]

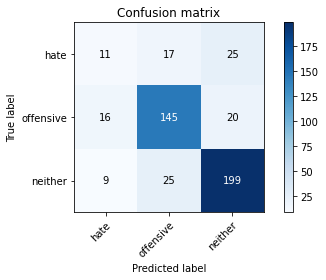
[ 9 25 199]]

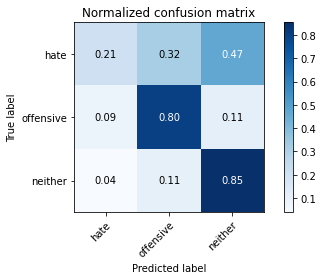
Normalized confusion matrix

[[0.20754717 0.32075472 0.47169811]

[0.08839779 0.80110497 0.11049724]

[0.03862661 0.10729614 0.85407725]]





precision recall f1-score support

hate 0.31 0.21 0.25 53

offensive 0.78 0.80 0.79 181

neither 0.82 0.85 0.83 233

accuracy 0.76 467

macro avg 0.63 0.62 0.62 467

weighted avg 0.74 0.76 0.75 467

average eval\_loss: 0.5856038040171067, accuracy: 0.7601713062098501

training 3 epoch...

172.15110206604004 seconds used......

3 training finished! average train loss: 0.714912659639404

total clf loss: 72.12397903203964 total adv loss: 39.003839794546366

evaluating...

Confusion matrix, without normalization

[[ 6 12 35]

[ 8 142 31]

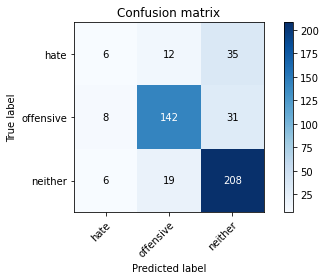
[ 6 19 208]]

Normalized confusion matrix

[[0.11320755 0.22641509 0.66037736]

[0.0441989 0.78453039 0.17127072]

[0.02575107 0.08154506 0.89270386]]





precision recall f1-score support

hate 0.30 0.11 0.16 53

offensive 0.82 0.78 0.80 181

neither 0.76 0.89 0.82 233

accuracy 0.76 467

macro avg 0.63 0.60 0.60 467

weighted avg 0.73 0.76 0.74 467

average eval\_loss: 0.6861196545263132, accuracy: 0.7623126338329764

training 4 epoch...

172.57683062553406 seconds used......

4 training finished! average train loss: 0.3602175803943759

total clf loss: 43.94604118168354 total adv loss: 15.849825578916352

evaluating...

Confusion matrix, without normalization

[[ 11 12 30]

[ 16 137 28]

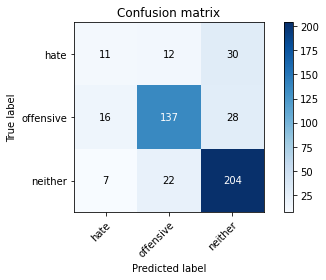
[ 7 22 204]]

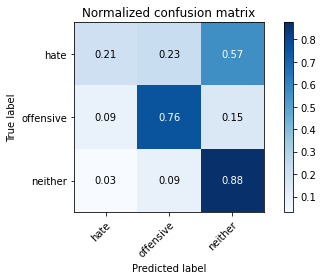
Normalized confusion matrix

[[0.20754717 0.22641509 0.56603774]

[0.08839779 0.75690608 0.15469613]

[0.03004292 0.0944206 0.87553648]]





precision recall f1-score support

hate 0.32 0.21 0.25 53

offensive 0.80 0.76 0.78 181

neither 0.78 0.88 0.82 233

accuracy 0.75 467

macro avg 0.63 0.61 0.62 467

weighted avg 0.74 0.75 0.74 467

average eval\_loss: 0.8037960318227609, accuracy: 0.7537473233404711

training 5 epoch...

172.54396963119507 seconds used......

5 training finished! average train loss: 0.22346551141568594

total clf loss: 29.137443479150534 total adv loss: 8.895156892685918

evaluating...

Confusion matrix, without normalization

[[ 13 10 30]

[ 18 138 25]

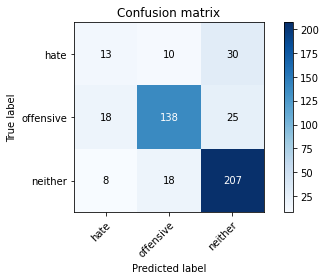
[ 8 18 207]]

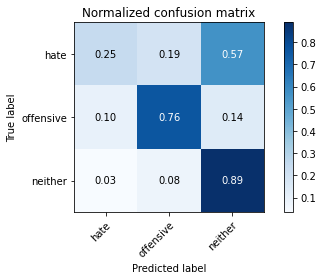
Normalized confusion matrix

[[0.24528302 0.18867925 0.56603774]

[0.09944751 0.76243094 0.13812155]

[0.03433476 0.07725322 0.88841202]]





precision recall f1-score support

hate 0.33 0.25 0.28 53

offensive 0.83 0.76 0.80 181

neither 0.79 0.89 0.84 233

accuracy 0.77 467

macro avg 0.65 0.63 0.64 467

weighted avg 0.75 0.77 0.76 467

average eval\_loss: 0.9767666719853878, accuracy: 0.7665952890792291

model of 5 epoch

Confusion matrix, without normalization

[[ 57 66 67]

[ 37 316 62]

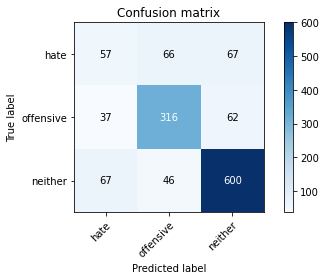
[ 67 46 600]]

Normalized confusion matrix

[[0.3 0.34736842 0.35263158]

[0.08915663 0.76144578 0.14939759]

[0.09396914 0.06451613 0.84151473]]





precision recall f1-score support

hate 0.35 0.30 0.32 190

offensive 0.74 0.76 0.75 415

neither 0.82 0.84 0.83 713

accuracy 0.74 1318

macro avg 0.64 0.63 0.64 1318

weighted avg 0.73 0.74 0.73 1318

average eval\_loss: 0.8924551590374021, accuracy: 0.7382397572078907

0.7382397572078907