hate as no-offensive

alpha=2

training 1 epoch...

26.28485870361328 seconds used......

1 training finished! average train loss: 2.20797630080155

total clf loss: 55.88251632452011 total adv loss: 33.88207855820656

evaluating...

Confusion matrix, without normalization

[[ 26 13 6]

[ 31 147 8]

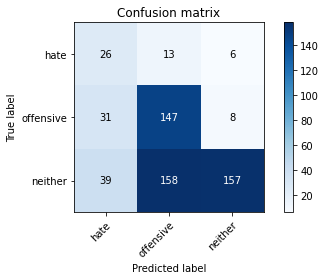
[ 39 158 157]]

Normalized confusion matrix

[[0.57777778 0.28888889 0.13333333]

[0.16666667 0.79032258 0.04301075]

[0.11016949 0.44632768 0.44350282]]





precision recall f1-score support

hate 0.27 0.58 0.37 45

offensive 0.46 0.79 0.58 186

neither 0.92 0.44 0.60 354

accuracy 0.56 585

macro avg 0.55 0.60 0.52 585

weighted avg 0.72 0.56 0.58 585

average eval\_loss: 0.8993544797102611, accuracy: 0.5641025641025641

training 2 epoch...

26.624375343322754 seconds used......

2 training finished! average train loss: 1.7258314809628896

total clf loss: 41.29577553272247 total adv loss: 27.675394028425217

evaluating...

Confusion matrix, without normalization

[[ 39 4 2]

[ 58 115 13]

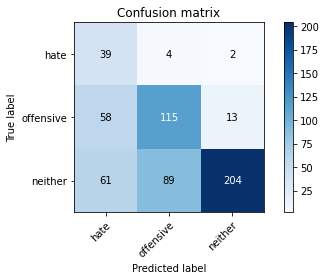
[ 61 89 204]]

Normalized confusion matrix

[[0.86666667 0.08888889 0.04444444]

[0.31182796 0.61827957 0.06989247]

[0.17231638 0.25141243 0.57627119]]





precision recall f1-score support

hate 0.25 0.87 0.38 45

offensive 0.55 0.62 0.58 186

neither 0.93 0.58 0.71 354

accuracy 0.61 585

macro avg 0.58 0.69 0.56 585

weighted avg 0.76 0.61 0.65 585

average eval\_loss: 0.8417076011498769, accuracy: 0.611965811965812

training 3 epoch...

27.07436466217041 seconds used......

3 training finished! average train loss: 1.0289663736309325

total clf loss: 25.304389387369156 total adv loss: 16.15886365622282

evaluating...

Confusion matrix, without normalization

[[ 33 6 6]

[ 33 107 46]

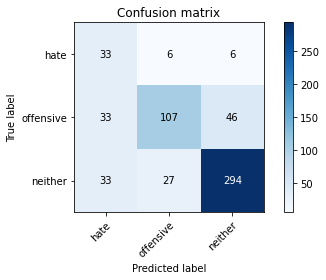
[ 33 27 294]]

Normalized confusion matrix

[[0.73333333 0.13333333 0.13333333]

[0.17741935 0.57526882 0.24731183]

[0.09322034 0.07627119 0.83050847]]





precision recall f1-score support

hate 0.33 0.73 0.46 45

offensive 0.76 0.58 0.66 186

neither 0.85 0.83 0.84 354

accuracy 0.74 585

macro avg 0.65 0.71 0.65 585

weighted avg 0.78 0.74 0.75 585

average eval\_loss: 0.6100472321112951, accuracy: 0.7418803418803419

training 4 epoch...

27.314727544784546 seconds used......

4 training finished! average train loss: 0.4148864960297942

total clf loss: 11.429073676466942 total adv loss: 5.902285106480122

evaluating...

Confusion matrix, without normalization

[[ 30 10 5]

[ 23 124 39]

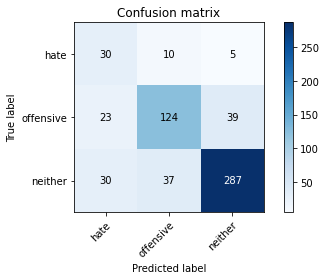
[ 30 37 287]]

Normalized confusion matrix

[[0.66666667 0.22222222 0.11111111]

[0.12365591 0.66666667 0.20967742]

[0.08474576 0.10451977 0.81073446]]





precision recall f1-score support

hate 0.36 0.67 0.47 45

offensive 0.73 0.67 0.69 186

neither 0.87 0.81 0.84 354

accuracy 0.75 585

macro avg 0.65 0.71 0.67 585

weighted avg 0.78 0.75 0.76 585

average eval\_loss: 0.6801939343412717, accuracy: 0.7538461538461538

training 5 epoch...

27.52495241165161 seconds used......

5 training finished! average train loss: 0.21084043077592338

total clf loss: 6.736049983650446 total adv loss: 2.5355070412624627

evaluating...

Confusion matrix, without normalization

[[ 35 4 6]

[ 40 111 35]

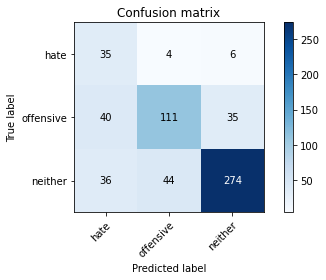
[ 36 44 274]]

Normalized confusion matrix

[[0.77777778 0.08888889 0.13333333]

[0.21505376 0.59677419 0.18817204]

[0.10169492 0.12429379 0.7740113 ]]





precision recall f1-score support

hate 0.32 0.78 0.45 45

offensive 0.70 0.60 0.64 186

neither 0.87 0.77 0.82 354

accuracy 0.72 585

macro avg 0.63 0.72 0.64 585

weighted avg 0.77 0.72 0.73 585

average eval\_loss: 0.8176290929317475, accuracy: 0.717948717948718

model of 2 epoch

Confusion matrix, without normalization

[[ 48 2 1]

[ 49 102 18]

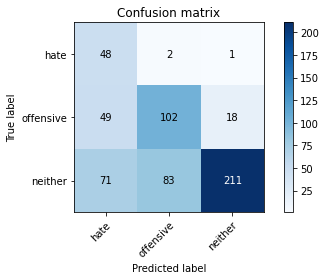
[ 71 83 211]]

Normalized confusion matrix

[[0.94117647 0.03921569 0.01960784]

[0.28994083 0.6035503 0.10650888]

[0.19452055 0.22739726 0.57808219]]





precision recall f1-score support

hate 0.29 0.94 0.44 51

offensive 0.55 0.60 0.57 169

neither 0.92 0.58 0.71 365

accuracy 0.62 585

macro avg 0.58 0.71 0.57 585

weighted avg 0.75 0.62 0.65 585

average eval\_loss: 0.9004030346870422, accuracy: 0.6170940170940171

0.6170940170940171