hate as no-offensive

alpha=2

training 1 epoch...

142.75095391273499 seconds used......

1 training finished! average train loss: 1.4234147040189895

total clf loss: 166.8455562889576 total adv loss: 77.42308314144611

evaluating...

Confusion matrix, without normalization

[[ 42 1 47]

[ 17 8 42]

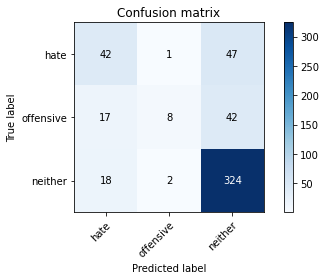
[ 18 2 324]]

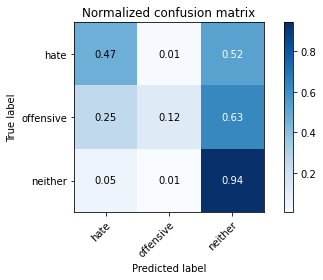
Normalized confusion matrix

[[0.46666667 0.01111111 0.52222222]

[0.25373134 0.11940299 0.62686567]

[0.05232558 0.00581395 0.94186047]]





precision recall f1-score support

hate 0.55 0.47 0.50 90

offensive 0.73 0.12 0.21 67

neither 0.78 0.94 0.86 344

accuracy 0.75 501

macro avg 0.69 0.51 0.52 501

weighted avg 0.73 0.75 0.71 501

average eval\_loss: 0.568450601055072, accuracy: 0.7465069860279441

training 2 epoch...

142.14071583747864 seconds used......

2 training finished! average train loss: 0.9829780174567636

total clf loss: 114.94937251508236 total adv loss: 53.601830122992396

evaluating...

Confusion matrix, without normalization

[[ 54 5 31]

[ 16 28 23]

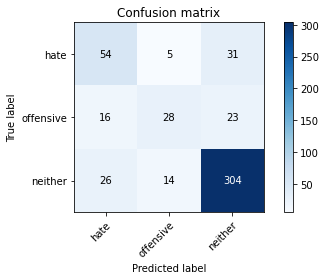
[ 26 14 304]]

Normalized confusion matrix

[[0.6 0.05555556 0.34444444]

[0.23880597 0.41791045 0.34328358]

[0.0755814 0.04069767 0.88372093]]





precision recall f1-score support

hate 0.56 0.60 0.58 90

offensive 0.60 0.42 0.49 67

neither 0.85 0.88 0.87 344

accuracy 0.77 501

macro avg 0.67 0.63 0.65 501

weighted avg 0.76 0.77 0.76 501

average eval\_loss: 0.6937602838644614, accuracy: 0.7704590818363274

training 3 epoch...

142.31170678138733 seconds used......

3 training finished! average train loss: 0.5974562737794049

total clf loss: 76.4595680013299 total adv loss: 29.28277505352162

evaluating...

Confusion matrix, without normalization

[[ 48 7 35]

[ 12 28 27]

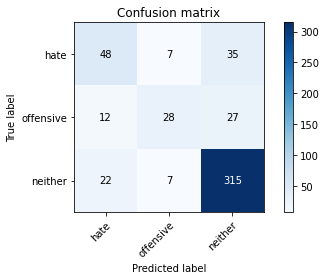
[ 22 7 315]]

Normalized confusion matrix

[[0.53333333 0.07777778 0.38888889]

[0.17910448 0.41791045 0.40298507]

[0.06395349 0.02034884 0.91569767]]





precision recall f1-score support

hate 0.59 0.53 0.56 90

offensive 0.67 0.42 0.51 67

neither 0.84 0.92 0.87 344

accuracy 0.78 501

macro avg 0.70 0.62 0.65 501

weighted avg 0.77 0.78 0.77 501

average eval\_loss: 0.6667453342905412, accuracy: 0.780439121756487

training 4 epoch...

142.0492343902588 seconds used......

4 training finished! average train loss: 0.32604470751784015

total clf loss: 49.11221024394035 total adv loss: 12.286946726362658

evaluating...

Confusion matrix, without normalization

[[ 49 1 40]

[ 23 18 26]

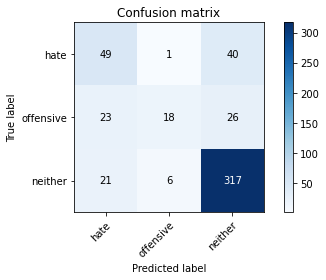
[ 21 6 317]]

Normalized confusion matrix

[[0.54444444 0.01111111 0.44444444]

[0.34328358 0.26865672 0.3880597 ]

[0.06104651 0.01744186 0.92151163]]





precision recall f1-score support

hate 0.53 0.54 0.54 90

offensive 0.72 0.27 0.39 67

neither 0.83 0.92 0.87 344

accuracy 0.77 501

macro avg 0.69 0.58 0.60 501

weighted avg 0.76 0.77 0.75 501

average eval\_loss: 0.7255283506730428, accuracy: 0.7664670658682635

training 5 epoch...

142.05121517181396 seconds used......

5 training finished! average train loss: 0.1722308157745386

total clf loss: 29.190394977107644 total adv loss: 4.8668846685140466

evaluating...

Confusion matrix, without normalization

[[ 47 6 37]

[ 15 25 27]

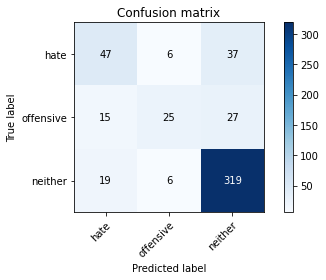
[ 19 6 319]]

Normalized confusion matrix

[[0.52222222 0.06666667 0.41111111]

[0.2238806 0.37313433 0.40298507]

[0.05523256 0.01744186 0.92732558]]





precision recall f1-score support

hate 0.58 0.52 0.55 90

offensive 0.68 0.37 0.48 67

neither 0.83 0.93 0.88 344

accuracy 0.78 501

macro avg 0.70 0.61 0.64 501

weighted avg 0.77 0.78 0.77 501

average eval\_loss: 0.9477990521834447, accuracy: 0.780439121756487

model of 4 epoch

Confusion matrix, without normalization

[[ 338 24 411]

[ 109 79 242]

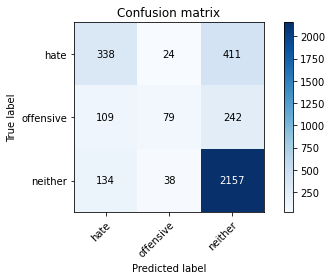
[ 134 38 2157]]

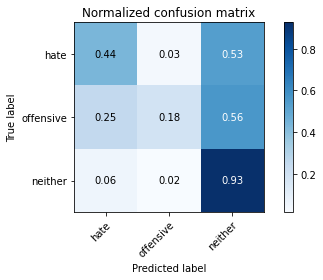
Normalized confusion matrix

[[0.43725744 0.03104787 0.5316947 ]

[0.25348837 0.18372093 0.5627907 ]

[0.05753542 0.01631602 0.92614856]]





precision recall f1-score support

hate 0.58 0.44 0.50 773

offensive 0.56 0.18 0.28 430

neither 0.77 0.93 0.84 2329

accuracy 0.73 3532

macro avg 0.64 0.52 0.54 3532

weighted avg 0.70 0.73 0.70 3532

average eval\_loss: 0.9161743291017026, accuracy: 0.7287655719139298

0.7287655719139298