hate as no-offensive

alpha=2

training 1 epoch...

836.2495894432068 seconds used......

1 training finished! average train loss: 0.7727744892284635

total clf loss: 294.4148911405355 total adv loss: 236.08870082721114

evaluating...

Confusion matrix, without normalization

[[ 29 79 26]

[ 28 1882 27]

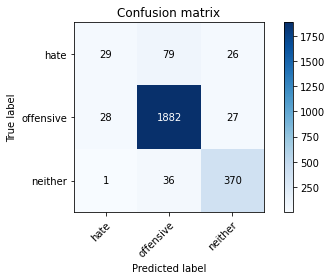
[ 1 36 370]]

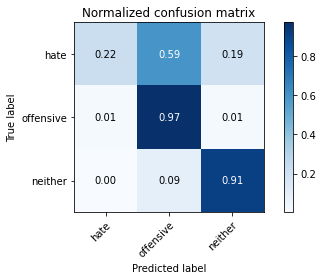
Normalized confusion matrix

[[0.21641791 0.58955224 0.19402985]

[0.01445534 0.97160558 0.01393908]

[0.002457 0.08845209 0.90909091]]





precision recall f1-score support

hate 0.50 0.22 0.30 134

offensive 0.94 0.97 0.96 1937

neither 0.87 0.91 0.89 407

accuracy 0.92 2478

macro avg 0.77 0.70 0.72 2478

weighted avg 0.91 0.92 0.91 2478

average eval\_loss: 0.24617940036489838, accuracy: 0.920500403551251

training 2 epoch...

835.7362463474274 seconds used......

2 training finished! average train loss: 0.5832751781638381

total clf loss: 212.52913575526327 total adv loss: 183.03992025461048

evaluating...

Confusion matrix, without normalization

[[ 44 77 13]

[ 33 1875 29]

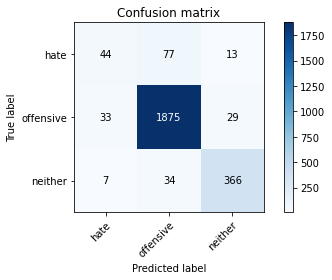
[ 7 34 366]]

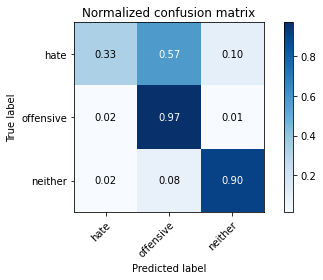
Normalized confusion matrix

[[0.32835821 0.57462687 0.09701493]

[0.01703665 0.96799174 0.01497161]

[0.01719902 0.08353808 0.8992629 ]]





precision recall f1-score support

hate 0.52 0.33 0.40 134

offensive 0.94 0.97 0.96 1937

neither 0.90 0.90 0.90 407

accuracy 0.92 2478

macro avg 0.79 0.73 0.75 2478

weighted avg 0.91 0.92 0.92 2478

average eval\_loss: 0.22567229455096588, accuracy: 0.9221146085552865

training 3 epoch...

836.0072870254517 seconds used......

3 training finished! average train loss: 0.4386675351359431

total clf loss: 159.15837691514753 total adv loss: 137.99990921863355

evaluating...

Confusion matrix, without normalization

[[ 35 81 18]

[ 20 1890 27]

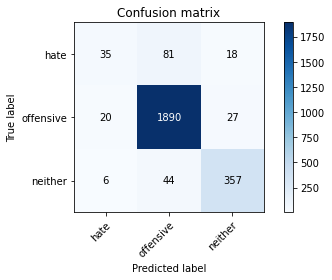
[ 6 44 357]]

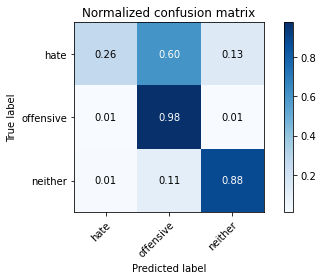
Normalized confusion matrix

[[0.26119403 0.60447761 0.13432836]

[0.01032525 0.97573567 0.01393908]

[0.01474201 0.10810811 0.87714988]]





precision recall f1-score support

hate 0.57 0.26 0.36 134

offensive 0.94 0.98 0.96 1937

neither 0.89 0.88 0.88 407

accuracy 0.92 2478

macro avg 0.80 0.70 0.73 2478

weighted avg 0.91 0.92 0.91 2478

average eval\_loss: 0.2511173191540424, accuracy: 0.9209039548022598

training 4 epoch...

835.7869024276733 seconds used......

4 training finished! average train loss: 0.3186629419287783

total clf loss: 114.63325331616215 total adv loss: 100.7401928534382

evaluating...

Confusion matrix, without normalization

[[ 50 72 12]

[ 37 1878 22]

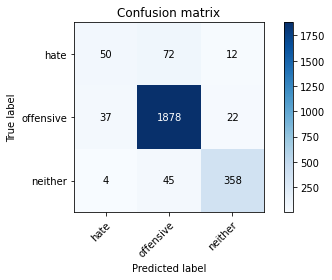
[ 4 45 358]]

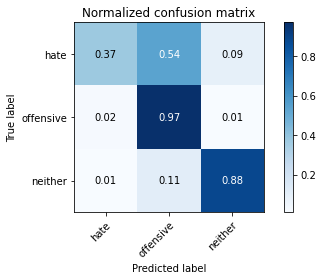
Normalized confusion matrix

[[0.37313433 0.53731343 0.08955224]

[0.0191017 0.96954053 0.01135777]

[0.00982801 0.11056511 0.87960688]]





precision recall f1-score support

hate 0.55 0.37 0.44 134

offensive 0.94 0.97 0.96 1937

neither 0.91 0.88 0.90 407

accuracy 0.92 2478

macro avg 0.80 0.74 0.77 2478

weighted avg 0.92 0.92 0.92 2478

average eval\_loss: 0.28047933648791046, accuracy: 0.9225181598062954

training 5 epoch...

836.2750360965729 seconds used......

5 training finished! average train loss: 0.19617922036488908

total clf loss: 70.73071439552587 total adv loss: 61.939536260390014

evaluating...

Confusion matrix, without normalization

[[ 46 73 15]

[ 49 1863 25]

[ 6 45 356]]

Normalized confusion matrix

[[0.34328358 0.54477612 0.1119403 ]

[0.02529685 0.96179659 0.01290656]

[0.01474201 0.11056511 0.87469287]]





precision recall f1-score support

hate 0.46 0.34 0.39 134

offensive 0.94 0.96 0.95 1937

neither 0.90 0.87 0.89 407

accuracy 0.91 2478

macro avg 0.76 0.73 0.74 2478

weighted avg 0.91 0.91 0.91 2478

average eval\_loss: 0.30069400856061085, accuracy: 0.914043583535109

model of 4 epoch

Confusion matrix, without normalization

[[ 55 64 5]

[ 63 1843 36]

[ 6 46 361]]

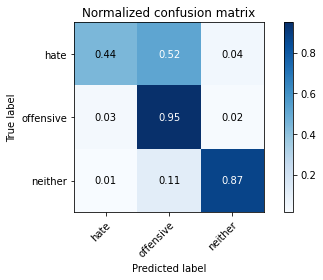
Normalized confusion matrix

[[0.44354839 0.51612903 0.04032258]

[0.03244078 0.94902163 0.01853759]

[0.01452785 0.11138015 0.87409201]]





precision recall f1-score support

hate 0.44 0.44 0.44 124

offensive 0.94 0.95 0.95 1942

neither 0.90 0.87 0.89 413

accuracy 0.91 2479

macro avg 0.76 0.76 0.76 2479

weighted avg 0.91 0.91 0.91 2479

average eval\_loss: 0.31009720439737243, accuracy: 0.9112545381202097