with adv

data without rebalance

training 1 epoch...

2263.594589471817 seconds used......

1 training finished! average train loss: 0.6757657190249635

total clf loss: 766.4677959587425 total adv loss: 488.165996751748

evaluating...

Confusion matrix, without normalization

[[ 350 328 314]

[ 155 5048 206]

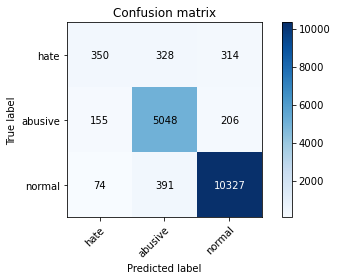
[ 74 391 10327]]

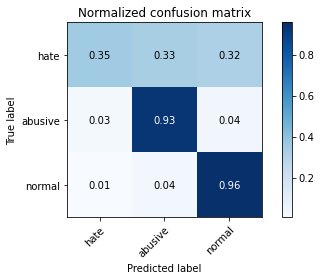
Normalized confusion matrix

[[0.35282258 0.33064516 0.31653226]

[0.02865594 0.93325938 0.03808467]

[0.00685693 0.03623054 0.95691253]]





precision recall f1-score support

hate 0.60 0.35 0.45 992

abusive 0.88 0.93 0.90 5409

normal 0.95 0.96 0.95 10792

accuracy 0.91 17193

macro avg 0.81 0.75 0.77 17193

weighted avg 0.91 0.91 0.91 17193

average eval\_loss: 0.2637084864296539, accuracy: 0.9146164136567208

training 2 epoch...

2263.836212873459 seconds used......

2 training finished! average train loss: 0.5462252981411875

total clf loss: 621.806493232958 total adv loss: 393.45427515683696

evaluating...

Confusion matrix, without normalization

[[ 264 372 356]

[ 90 5003 316]

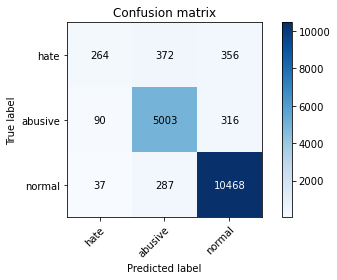
[ 37 287 10468]]

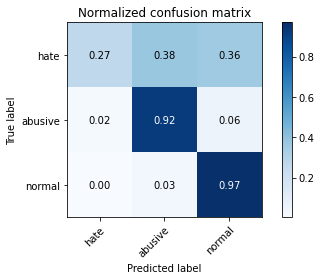
Normalized confusion matrix

[[0.26612903 0.375 0.35887097]

[0.01663894 0.92493991 0.05842115]

[0.00342847 0.02659377 0.96997776]]





precision recall f1-score support

hate 0.68 0.27 0.38 992

abusive 0.88 0.92 0.90 5409

normal 0.94 0.97 0.95 10792

accuracy 0.92 17193

macro avg 0.83 0.72 0.75 17193

weighted avg 0.91 0.92 0.91 17193

average eval\_loss: 0.27448123330250385, accuracy: 0.9151980457162798

training 3 epoch...

2262.257097005844 seconds used......

3 training finished! average train loss: 0.4301450678673524

total clf loss: 499.3572846390307 total adv loss: 304.99342281247664

evaluating...

Confusion matrix, without normalization

[[ 344 380 268]

[ 120 5058 231]

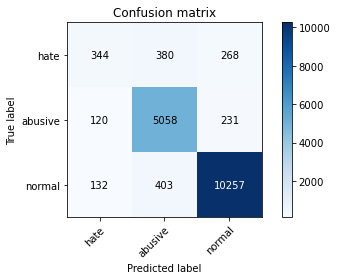
[ 132 403 10257]]

Normalized confusion matrix

[[0.34677419 0.38306452 0.27016129]

[0.02218525 0.93510815 0.0427066 ]

[0.01223128 0.03734248 0.95042624]]





precision recall f1-score support

hate 0.58 0.35 0.43 992

abusive 0.87 0.94 0.90 5409

normal 0.95 0.95 0.95 10792

accuracy 0.91 17193

macro avg 0.80 0.74 0.76 17193

weighted avg 0.90 0.91 0.91 17193

average eval\_loss: 0.29211299093223586, accuracy: 0.9107776420636305

training 4 epoch...

2261.695606470108 seconds used......

4 training finished! average train loss: 0.33139946766576134

total clf loss: 386.8204717049375 total adv loss: 233.92937709781108

evaluating...

Confusion matrix, without normalization

[[ 331 396 265]

[ 137 5044 228]

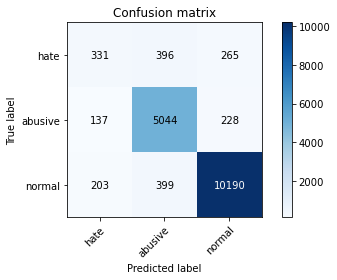
[ 203 399 10190]]

Normalized confusion matrix

[[0.33366935 0.39919355 0.2671371 ]

[0.02532816 0.93251987 0.04215197]

[0.01881023 0.03697183 0.94421794]]





precision recall f1-score support

hate 0.49 0.33 0.40 992

abusive 0.86 0.93 0.90 5409

normal 0.95 0.94 0.95 10792

accuracy 0.91 17193

macro avg 0.77 0.74 0.75 17193

weighted avg 0.90 0.91 0.90 17193

average eval\_loss: 0.32766420661233536, accuracy: 0.9053103007037748

training 5 epoch...

2263.193878889084 seconds used......

5 training finished! average train loss: 0.2514299264644605

total clf loss: 290.11591309378855 total adv loss: 179.1609333033848

evaluating...

Confusion matrix, without normalization

[[ 316 424 252]

[ 119 5085 205]

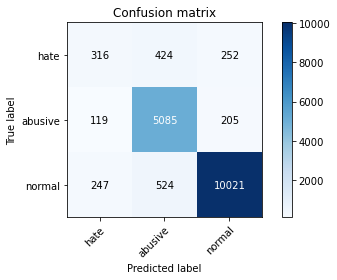
[ 247 524 10021]]

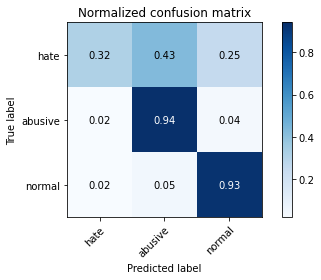
Normalized confusion matrix

[[0.31854839 0.42741935 0.25403226]

[0.02200037 0.94009983 0.0378998 ]

[0.02288732 0.04855448 0.92855819]]





precision recall f1-score support

hate 0.46 0.32 0.38 992

abusive 0.84 0.94 0.89 5409

normal 0.96 0.93 0.94 10792

accuracy 0.90 17193

macro avg 0.75 0.73 0.74 17193

weighted avg 0.89 0.90 0.89 17193

average eval\_loss: 0.3694932456507326, accuracy: 0.8969929622520794