hate as offensive

alpha=0.05

1 run...

Downloading: 100%

411/411 [00:00<00:00, 13.8kB/s]

Downloading: 100%

263M/263M [00:12<00:00, 16.5MB/s]

training 1 epoch...

1317.6059482097626 seconds used......

1 training finished! average train loss: 0.3733054033652647

total clf loss: 525.6976101733744 total adv loss: 330.1807906962931

evaluating...

Confusion matrix, without normalization

[[ 322 396 304]

[ 111 5085 211]

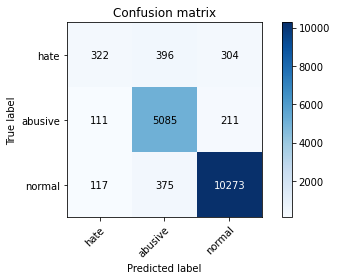
[ 117 375 10273]]

Normalized confusion matrix

[[0.31506849 0.38747554 0.29745597]

[0.02052894 0.94044757 0.03902349]

[0.01086856 0.03483511 0.95429633]]





precision recall f1-score support

hate 0.59 0.32 0.41 1022

abusive 0.87 0.94 0.90 5407

normal 0.95 0.95 0.95 10765

accuracy 0.91 17194

macro avg 0.80 0.74 0.76 17194

weighted avg 0.90 0.91 0.91 17194

average eval\_loss: 0.2741241004772831, accuracy: 0.9119460276840758

training 2 epoch...

1320.7905673980713 seconds used......

2 training finished! average train loss: 0.2856717638253532

total clf loss: 401.7724743243307 total adv loss: 242.32377092447132

evaluating...

Confusion matrix, without normalization

[[ 487 220 315]

[ 462 4620 325]

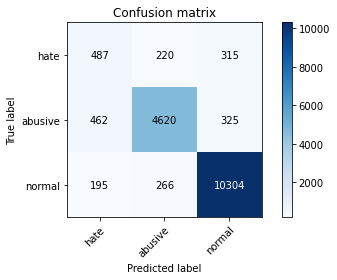
[ 195 266 10304]]

Normalized confusion matrix

[[0.47651663 0.21526419 0.30821918]

[0.08544479 0.85444794 0.06010727]

[0.01811426 0.02470971 0.95717603]]





precision recall f1-score support

hate 0.43 0.48 0.45 1022

abusive 0.90 0.85 0.88 5407

normal 0.94 0.96 0.95 10765

accuracy 0.90 17194

macro avg 0.76 0.76 0.76 17194

weighted avg 0.90 0.90 0.90 17194

average eval\_loss: 0.2990489646792412, accuracy: 0.8963010352448528

training 3 epoch...

1320.8053803443909 seconds used......

3 training finished! average train loss: 0.2119530036941695

total clf loss: 297.4170517390594 total adv loss: 166.26309876097366

evaluating...

Confusion matrix, without normalization

[[ 396 358 268]

[ 222 4970 215]

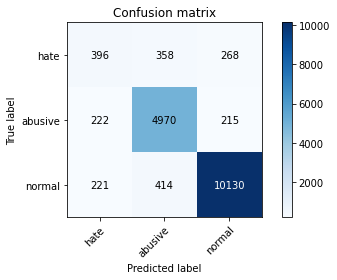
[ 221 414 10130]]

Normalized confusion matrix

[[0.38747554 0.35029354 0.26223092]

[0.04105789 0.91917884 0.03976327]

[0.02052949 0.03845797 0.94101254]]





precision recall f1-score support

hate 0.47 0.39 0.43 1022

abusive 0.87 0.92 0.89 5407

normal 0.95 0.94 0.95 10765

accuracy 0.90 17194

macro avg 0.76 0.75 0.75 17194

weighted avg 0.90 0.90 0.90 17194

average eval\_loss: 0.3164738463508615, accuracy: 0.9012446202163545

training 4 epoch...

1319.8371675014496 seconds used......

4 training finished! average train loss: 0.14149056304443514

total clf loss: 198.23940046411008 total adv loss: 104.92544903361704

evaluating...

Confusion matrix, without normalization

[[ 495 309 218]

[ 390 4851 166]

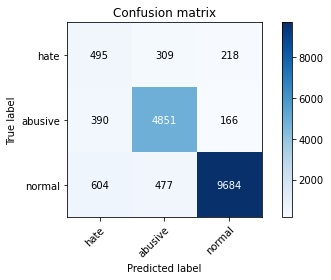
[ 604 477 9684]]

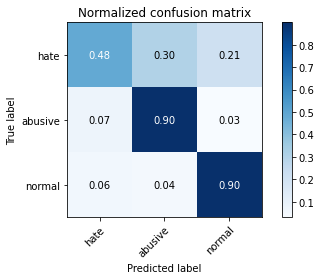
Normalized confusion matrix

[[0.48434442 0.30234834 0.21330724]

[0.07212872 0.89717033 0.03070094]

[0.05610776 0.04431026 0.89958198]]





precision recall f1-score support

hate 0.33 0.48 0.39 1022

abusive 0.86 0.90 0.88 5407

normal 0.96 0.90 0.93 10765

accuracy 0.87 17194

macro avg 0.72 0.76 0.73 17194

weighted avg 0.89 0.87 0.88 17194

average eval\_loss: 0.4567728662874203, accuracy: 0.8741421426078865

training 5 epoch...

1318.8542284965515 seconds used......

5 training finished! average train loss: 0.09873050965998278

total clf loss: 138.13810955709778 total adv loss: 69.39388723653974

evaluating...

Confusion matrix, without normalization

[[ 390 325 307]

[ 301 4840 266]

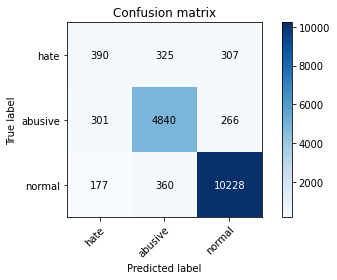
[ 177 360 10228]]

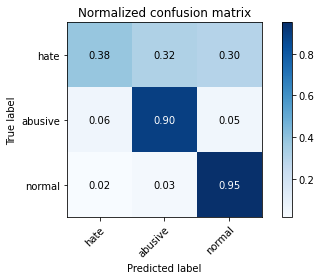
Normalized confusion matrix

[[0.3816047 0.31800391 0.30039139]

[0.05566858 0.89513593 0.04919549]

[0.01644217 0.03344171 0.95011612]]





precision recall f1-score support

hate 0.45 0.38 0.41 1022

abusive 0.88 0.90 0.89 5407

normal 0.95 0.95 0.95 10765

accuracy 0.90 17194

macro avg 0.76 0.74 0.75 17194

weighted avg 0.90 0.90 0.90 17194

average eval\_loss: 0.4526562531568045, accuracy: 0.8990345469349773

2 run...

training 1 epoch...

1320.0573000907898 seconds used......

1 training finished! average train loss: 0.37960553726951574

total clf loss: 534.7494508977979 total adv loss: 339.34997306670994

evaluating...

Confusion matrix, without normalization

[[ 299 376 347]

[ 116 5028 263]

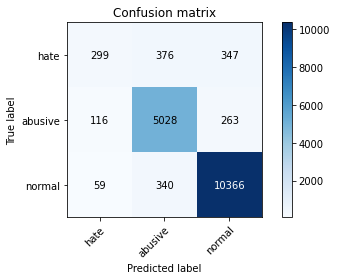
[ 59 340 10366]]

Normalized confusion matrix

[[0.2925636 0.36790607 0.33953033]

[0.02145367 0.92990568 0.04864065]

[0.00548072 0.03158384 0.96293544]]





precision recall f1-score support

hate 0.63 0.29 0.40 1022

abusive 0.88 0.93 0.90 5407

normal 0.94 0.96 0.95 10765

accuracy 0.91 17194

macro avg 0.82 0.73 0.75 17194

weighted avg 0.90 0.91 0.90 17194

average eval\_loss: 0.2787478981084775, accuracy: 0.9127021053855996

training 2 epoch...

1320.6013352870941 seconds used......

2 training finished! average train loss: 0.29129828906773736

total clf loss: 409.7652708683163 total adv loss: 248.6880910382606

evaluating...

Confusion matrix, without normalization

[[ 429 322 271]

[ 200 4968 239]

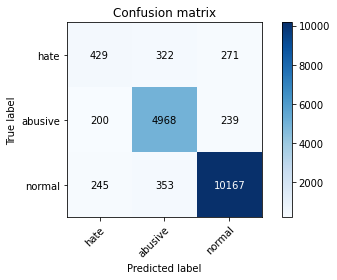
[ 245 353 10167]]

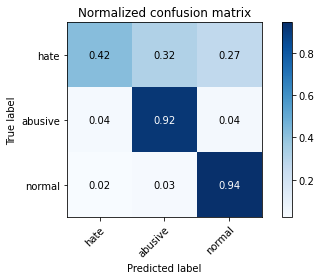
Normalized confusion matrix

[[0.41976517 0.31506849 0.26516634]

[0.03698909 0.91880895 0.04420196]

[0.02275894 0.03279145 0.94444961]]





precision recall f1-score support

hate 0.49 0.42 0.45 1022

abusive 0.88 0.92 0.90 5407

normal 0.95 0.94 0.95 10765

accuracy 0.91 17194

macro avg 0.77 0.76 0.77 17194

weighted avg 0.90 0.91 0.90 17194

average eval\_loss: 0.28650525304553814, accuracy: 0.9051994881935559

training 3 epoch...

1321.705950975418 seconds used......

3 training finished! average train loss: 0.21409514789739925

total clf loss: 300.5075999982655 total adv loss: 169.6363707734272

evaluating...

Confusion matrix, without normalization

[[ 373 291 358]

[ 290 4784 333]

[ 92 301 10372]]

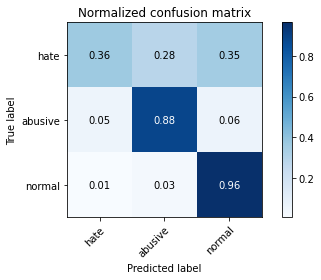
Normalized confusion matrix

[[0.36497065 0.28473581 0.35029354]

[0.05363418 0.88477899 0.06158683]

[0.00854621 0.02796098 0.9634928 ]]





precision recall f1-score support

hate 0.49 0.36 0.42 1022

abusive 0.89 0.88 0.89 5407

normal 0.94 0.96 0.95 10765

accuracy 0.90 17194

macro avg 0.77 0.74 0.75 17194

weighted avg 0.90 0.90 0.90 17194

average eval\_loss: 0.34320499380660613, accuracy: 0.903163894381761

training 4 epoch...

1320.6008014678955 seconds used......

4 training finished! average train loss: 0.14149233900233174

total clf loss: 198.12684740358964 total adv loss: 102.62593631213531

evaluating...

Confusion matrix, without normalization

[[ 508 291 223]

[ 397 4824 186]

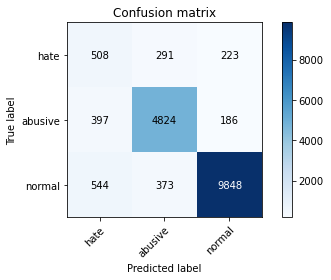
[ 544 373 9848]]

Normalized confusion matrix

[[0.49706458 0.28473581 0.21819961]

[0.07342334 0.89217681 0.03439985]

[0.05053414 0.03464933 0.91481654]]





1321.035269498825 seconds used......

5 training finished! average train loss: 0.09826731375354421

total clf loss: 137.583431328414 total adv loss: 70.93630725389812

evaluating...

Confusion matrix, without normalization

[[ 454 309 259]

[ 346 4804 257]

[ 383 366 10016]]

Normalized confusion matrix

[[0.44422701 0.30234834 0.25342466]

[0.06399112 0.8884779 0.04753098]

[0.03557826 0.03399907 0.93042267]]





precision recall f1-score support

hate 0.38 0.44 0.41 1022

abusive 0.88 0.89 0.88 5407

normal 0.95 0.93 0.94 10765

accuracy 0.89 17194

macro avg 0.74 0.75 0.74 17194

weighted avg 0.89 0.89 0.89 17194

average eval\_loss: 0.48005566010928474, accuracy: 0.888333139467256