input for adv: mean of all 7 layer

1 run...

Downloading: 100%

411/411 [00:00<00:00, 10.8kB/s]

Downloading: 100%

263M/263M [00:09<00:00, 16.3MB/s]

training 1 epoch...

1222.9644131660461 seconds used......

1 training finished! average train loss: 0.3695047152531811

total clf loss: 521.0523888841271 total adv loss: 340.9591492563486

evaluating...

Confusion matrix, without normalization

[[ 446 316 260]

[ 219 4951 237]

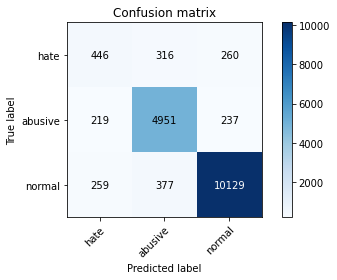
[ 259 377 10129]]

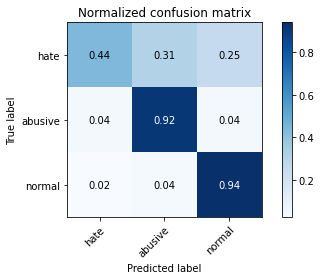
Normalized confusion matrix

[[0.43639922 0.30919765 0.25440313]

[0.04050305 0.91566488 0.04383207]

[0.02405945 0.0350209 0.94091965]]





precision recall f1-score support

hate 0.48 0.44 0.46 1022

abusive 0.88 0.92 0.90 5407

normal 0.95 0.94 0.95 10765

accuracy 0.90 17194

macro avg 0.77 0.76 0.77 17194

weighted avg 0.90 0.90 0.90 17194

average eval\_loss: 0.2955719477648652, accuracy: 0.9029894149121787

training 2 epoch...

1221.7121138572693 seconds used......

2 training finished! average train loss: 0.28546969251450466

total clf loss: 402.0648989379406 total adv loss: 253.68476128950715

evaluating...

Confusion matrix, without normalization

[[ 365 353 304]

[ 160 5016 231]

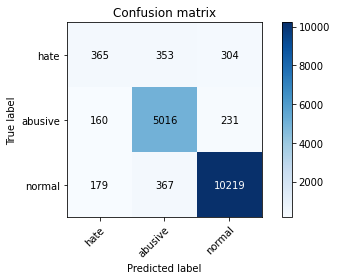
[ 179 367 10219]]

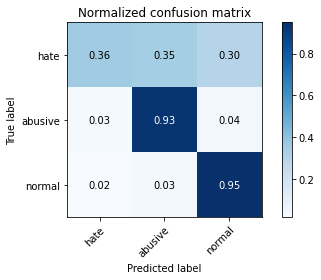
Normalized confusion matrix

[[0.35714286 0.34540117 0.29745597]

[0.02959127 0.92768633 0.0427224 ]

[0.01662796 0.03409196 0.94928007]]





precision recall f1-score support

hate 0.52 0.36 0.42 1022

abusive 0.87 0.93 0.90 5407

normal 0.95 0.95 0.95 10765

accuracy 0.91 17194

macro avg 0.78 0.74 0.76 17194

weighted avg 0.90 0.91 0.90 17194

average eval\_loss: 0.2853124908041642, accuracy: 0.9072932418285449

training 3 epoch...

1221.3169441223145 seconds used......

3 training finished! average train loss: 0.21144674743165173

total clf loss: 297.56403529830277 total adv loss: 183.01343492837623

evaluating...

Confusion matrix, without normalization

[[ 424 318 280]

[ 290 4883 234]

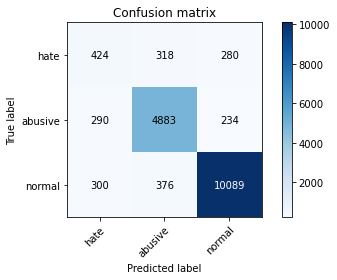
[ 300 376 10089]]

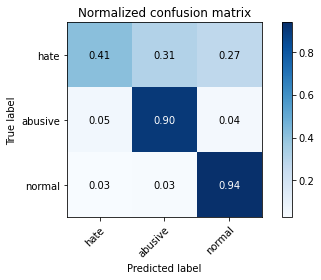
Normalized confusion matrix

[[0.4148728 0.3111546 0.2739726 ]

[0.05363418 0.90308859 0.04327723]

[0.02786809 0.03492801 0.9372039 ]]





precision recall f1-score support

hate 0.42 0.41 0.42 1022

abusive 0.88 0.90 0.89 5407

normal 0.95 0.94 0.94 10765

accuracy 0.90 17194

macro avg 0.75 0.75 0.75 17194

weighted avg 0.90 0.90 0.90 17194

average eval\_loss: 0.31645288993063014, accuracy: 0.8954286378969408

training 4 epoch...

1 run...

Downloading: 100%

411/411 [00:00<00:00, 14.3kB/s]

Downloading: 100%

263M/263M [00:06<00:00, 43.5MB/s]

training 1 epoch...

1201.5861785411835 seconds used......

1 training finished! average train loss: 0.3714678262434135

total clf loss: 523.9457078240812 total adv loss: 345.27185460552573

evaluating...

Confusion matrix, without normalization

[[ 375 349 298]

[ 162 5003 242]

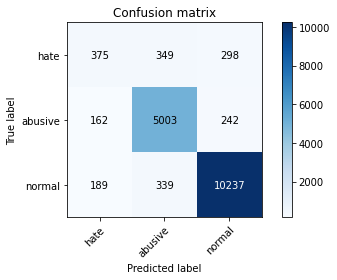
[ 189 339 10237]]

Normalized confusion matrix

[[0.36692759 0.34148728 0.29158513]

[0.02996116 0.92528204 0.0447568 ]

[0.0175569 0.03149094 0.95095216]]





precision recall f1-score support

hate 0.52 0.37 0.43 1022

abusive 0.88 0.93 0.90 5407

normal 0.95 0.95 0.95 10765

accuracy 0.91 17194

macro avg 0.78 0.75 0.76 17194

weighted avg 0.90 0.91 0.90 17194

average eval\_loss: 0.2814394134378364, accuracy: 0.9081656391764569

training 2 epoch...

1201.6188254356384 seconds used......

2 training finished! average train loss: 0.28617420373894165

total clf loss: 403.2392900120467 total adv loss: 257.9535263264552

evaluating...

Confusion matrix, without normalization

[[ 510 249 263]

[ 408 4779 220]

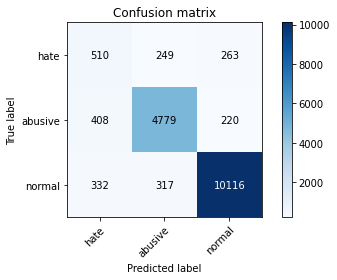
[ 332 317 10116]]

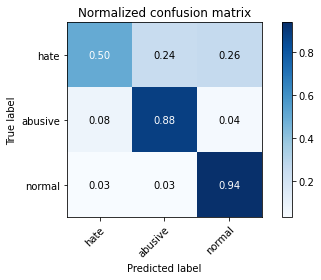
Normalized confusion matrix

[[0.49902153 0.24363992 0.25733855]

[0.07545774 0.88385426 0.040688 ]

[0.03084069 0.02944728 0.93971203]]





precision recall f1-score support

hate 0.41 0.50 0.45 1022

abusive 0.89 0.88 0.89 5407

normal 0.95 0.94 0.95 10765

accuracy 0.90 17194

macro avg 0.75 0.77 0.76 17194

weighted avg 0.90 0.90 0.90 17194

average eval\_loss: 0.2971437275626285, accuracy: 0.8959520763056881

training 3 epoch...

1201.3850390911102 seconds used......

3 training finished! average train loss: 0.2135949914659103

total clf loss: 300.68736639712006 total adv loss: 186.8759674578905

evaluating...

Confusion matrix, without normalization

[[ 419 321 282]

[ 264 4893 250]

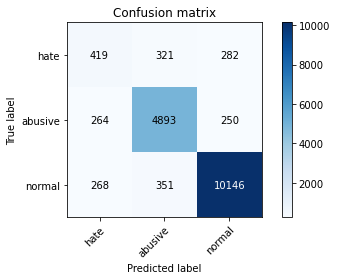
[ 268 351 10146]]

Normalized confusion matrix

[[0.40998043 0.31409002 0.27592955]

[0.0488256 0.90493804 0.04623636]

[0.02489549 0.03260567 0.94249884]]





precision recall f1-score support

hate 0.44 0.41 0.42 1022

abusive 0.88 0.90 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: 0.3262445167919924, accuracy: 0.8990345469349773

training 4 epoch...

1201.9445023536682 seconds used......

4 training finished! average train loss: 0.14194761803547806

total clf loss: 199.1812559706159 total adv loss: 111.2941012212541

evaluating...

Confusion matrix, without normalization

[[ 351 368 303]

[ 194 4992 221]

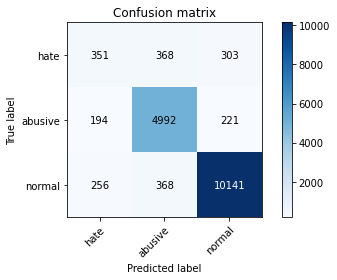
[ 256 368 10141]]

Normalized confusion matrix

[[0.34344423 0.36007828 0.2964775 ]

[0.03587942 0.92324764 0.04087294]

[0.02378077 0.03418486 0.94203437]]





precision recall f1-score support

hate 0.44 0.34 0.39 1022

abusive 0.87 0.92 0.90 5407

normal 0.95 0.94 0.95 10765

accuracy 0.90 17194

macro avg 0.75 0.74 0.74 17194

weighted avg 0.90 0.90 0.90 17194

average eval\_loss: 0.399576231523222, accuracy: 0.9005467023380249

training 5 epoch...

1201.0789813995361 seconds used......

5 training finished! average train loss: 0.10029962065454184

total clf loss: 140.79570326267276 total adv loss: 79.74041241614032

evaluating...

Confusion matrix, without normalization

[[ 479 296 247]

[ 407 4762 238]

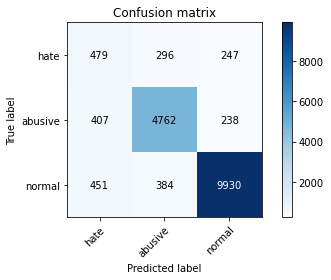
[ 451 384 9930]]

Normalized confusion matrix

[[0.46868885 0.28962818 0.24168297]

[0.07527279 0.88071019 0.04401701]

[0.04189503 0.03567116 0.92243381]]





precision recall f1-score support

hate 0.36 0.47 0.41 1022

abusive 0.88 0.88 0.88 5407

normal 0.95 0.92 0.94 10765

accuracy 0.88 17194

macro avg 0.73 0.76 0.74 17194

weighted avg 0.89 0.88 0.89 17194

average eval\_loss: 0.472729744035534, accuracy: 0.8823426776782599

2 run...

training 1 epoch...

1201.508226633072 seconds used......

1 training finished! average train loss: 0.3693637551488546

total clf loss: 520.911043625325 total adv loss: 341.97763039916754

evaluating...

Confusion matrix, without normalization

[[ 352 310 360]

[ 149 4918 340]

[ 72 280 10413]]

Normalized confusion matrix

[[0.3444227 0.30332681 0.35225049]

[0.02755687 0.90956168 0.06288145]

[0.00668834 0.02601022 0.96730144]]





precision recall f1-score support

hate 0.61 0.34 0.44 1022

abusive 0.89 0.91 0.90 5407

normal 0.94 0.97 0.95 10765

accuracy 0.91 17194

macro avg 0.81 0.74 0.76 17194

weighted avg 0.90 0.91 0.91 17194

average eval\_loss: 0.27655611152565757, accuracy: 0.9121205071536582

training 2 epoch...

1201.373556137085 seconds used......

2 training finished! average train loss: 0.288570935540229

total clf loss: 406.3829933684319 total adv loss: 255.44474667450413

evaluating...

Confusion matrix, without normalization

[[ 372 316 334]

[ 206 4902 299]

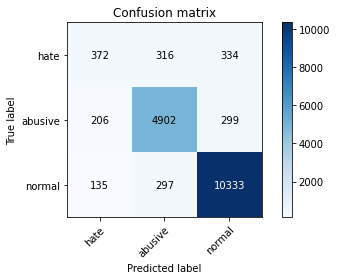
[ 135 297 10333]]

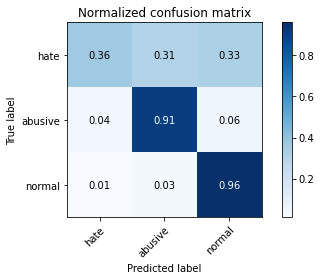
Normalized confusion matrix

[[0.36399217 0.30919765 0.32681018]

[0.03809876 0.90660255 0.05529869]

[0.01254064 0.02758941 0.95986995]]





precision recall f1-score support

hate 0.52 0.36 0.43 1022

abusive 0.89 0.91 0.90 5407

normal 0.94 0.96 0.95 10765

accuracy 0.91 17194

macro avg 0.78 0.74 0.76 17194

weighted avg 0.90 0.91 0.90 17194

average eval\_loss: 0.28271390214605735, accuracy: 0.9077003605909038

training 3 epoch...

1201.170163154602 seconds used......

3 training finished! average train loss: 0.21029734093753233

total clf loss: 295.80219655670226 total adv loss: 179.13246854115278

evaluating...

Confusion matrix, without normalization

[[ 519 245 258]

[ 436 4764 207]

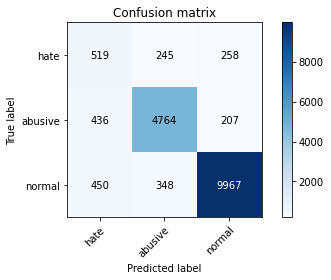
[ 450 348 9967]]

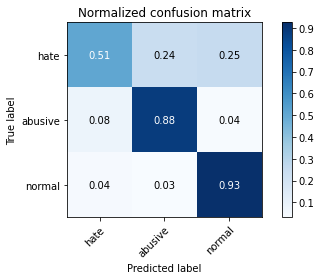
Normalized confusion matrix

[[0.50782779 0.23972603 0.25244618]

[0.08063621 0.88108008 0.03828371]

[0.04180214 0.03232699 0.92587088]]





precision recall f1-score support

hate 0.37 0.51 0.43 1022

abusive 0.89 0.88 0.89 5407

normal 0.96 0.93 0.94 10765

accuracy 0.89 17194

macro avg 0.74 0.77 0.75 17194

weighted avg 0.90 0.89 0.89 17194

average eval\_loss: 0.35163345974761734, accuracy: 0.8869373037105968

training 4 epoch...

1201.325667142868 seconds used......

4 training finished! average train loss: 0.142126186208525

total clf loss: 199.3989780433476 total adv loss: 110.7771990657784

evaluating...

Confusion matrix, without normalization

[[ 502 297 223]

[ 369 4829 209]

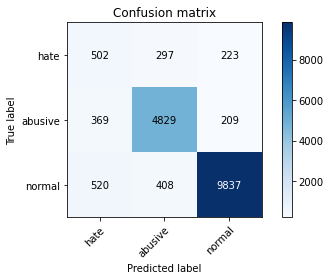
[ 520 408 9837]]

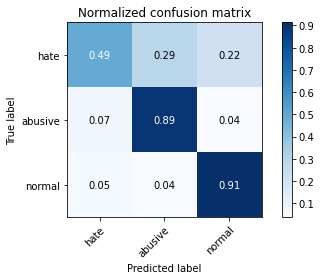
Normalized confusion matrix

[[0.49119374 0.29060665 0.21819961]

[0.06824487 0.89310154 0.0386536 ]

[0.04830469 0.0379006 0.91379471]]





precision recall f1-score support

hate 0.36 0.49 0.42 1022

abusive 0.87 0.89 0.88 5407

normal 0.96 0.91 0.94 10765

accuracy 0.88 17194

macro avg 0.73 0.77 0.74 17194

weighted avg 0.90 0.88 0.89 17194

average eval\_loss: 0.42729779148994146, accuracy: 0.8821681982086774

training 5 epoch...

1201.9799282550812 seconds used......

5 training finished! average train loss: 0.09986870040882534

total clf loss: 140.121308351052 total adv loss: 78.00802068959456

evaluating...

Confusion matrix, without normalization

[[ 462 260 300]

[ 483 4635 289]

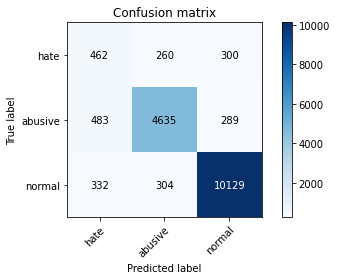
[ 332 304 10129]]

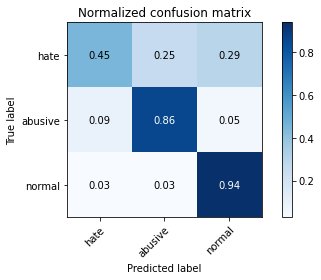
Normalized confusion matrix

[[0.45205479 0.25440313 0.29354207]

[0.08932865 0.85722212 0.05344923]

[0.03084069 0.02823967 0.94091965]]





precision recall f1-score support

hate 0.36 0.45 0.40 1022

abusive 0.89 0.86 0.87 5407

normal 0.95 0.94 0.94 10765

accuracy 0.89 17194

macro avg 0.73 0.75 0.74 17194

weighted avg 0.89 0.89 0.89 17194

average eval\_loss: 0.46955153277187156, accuracy: 0.885541467953937