input for adv: mean of last 2 layer

1 run...

Downloading: 100%

411/411 [00:00<00:00, 14.9kB/s]

Downloading: 100%

263M/263M [00:08<00:00, 32.4MB/s]

training 1 epoch...

1200.6712806224823 seconds used......

1 training finished! average train loss: 0.37005103017883567

total clf loss: 521.2184095289558 total adv loss: 329.37610072270036

evaluating...

Confusion matrix, without normalization

[[ 528 225 269]

[ 448 4683 276]

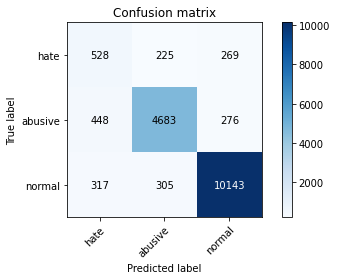
[ 317 305 10143]]

Normalized confusion matrix

[[0.51663405 0.22015656 0.26320939]

[0.08285556 0.8660995 0.05104494]

[0.02944728 0.02833256 0.94222016]]





precision recall f1-score support

hate 0.41 0.52 0.46 1022

abusive 0.90 0.87 0.88 5407

normal 0.95 0.94 0.95 10765

accuracy 0.89 17194

macro avg 0.75 0.77 0.76 17194

weighted avg 0.90 0.89 0.90 17194

average eval\_loss: 0.30709538547278836, accuracy: 0.892985925322787

training 2 epoch...

1200.1033766269684 seconds used......

2 training finished! average train loss: 0.28577227174110514

total clf loss: 402.1138117723167 total adv loss: 246.40865180734545

evaluating...

Confusion matrix, without normalization

[[ 381 326 315]

[ 229 4879 299]

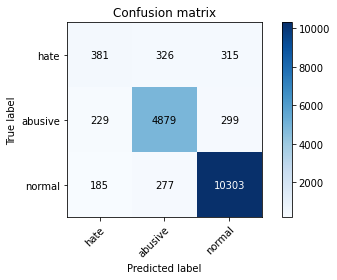
[ 185 277 10303]]

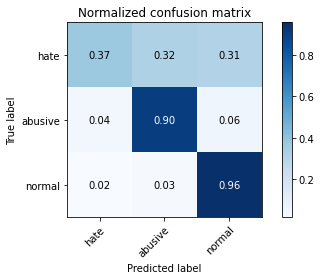
Normalized confusion matrix

[[0.37279843 0.31898239 0.30821918]

[0.04235251 0.90234881 0.05529869]

[0.01718532 0.02573154 0.95708314]]





precision recall f1-score support

hate 0.48 0.37 0.42 1022

abusive 0.89 0.90 0.90 5407

normal 0.94 0.96 0.95 10765

accuracy 0.91 17194

macro avg 0.77 0.74 0.76 17194

weighted avg 0.90 0.91 0.90 17194

average eval\_loss: 0.2892199426304636, accuracy: 0.9051413283703618

training 3 epoch...

1200.2024989128113 seconds used......

3 training finished! average train loss: 0.21011192178629515

total clf loss: 294.9392421236262 total adv loss: 166.93161166459322

evaluating...

Confusion matrix, without normalization

[[ 397 322 303]

[ 268 4818 321]

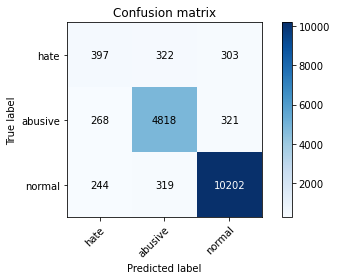
[ 244 319 10202]]

Normalized confusion matrix

[[0.38845401 0.31506849 0.2964775 ]

[0.04956538 0.89106714 0.05936749]

[0.02266605 0.02963307 0.94770088]]





precision recall f1-score support

hate 0.44 0.39 0.41 1022

abusive 0.88 0.89 0.89 5407

normal 0.94 0.95 0.95 10765

accuracy 0.90 17194

macro avg 0.75 0.74 0.75 17194

weighted avg 0.89 0.90 0.89 17194

average eval\_loss: 0.3310483479742394, accuracy: 0.8966499941840177

training 4 epoch...

1200.9270372390747 seconds used......

4 training finished! average train loss: 0.143252799534949

total clf loss: 200.74418997950852 total adv loss: 106.94742875371594

evaluating...

Confusion matrix, without normalization

[[ 447 314 261]

[ 341 4851 215]

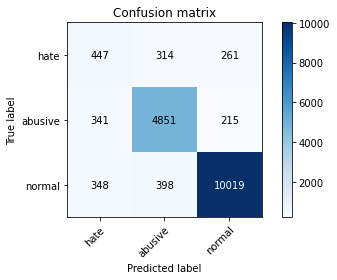
[ 348 398 10019]]

Normalized confusion matrix

[[0.43737769 0.3072407 0.2553816 ]

[0.0630664 0.89717033 0.03976327]

[0.03232699 0.03697167 0.93070135]]





precision recall f1-score support

hate 0.39 0.44 0.41 1022

abusive 0.87 0.90 0.88 5407

normal 0.95 0.93 0.94 10765

accuracy 0.89 17194

macro avg 0.74 0.76 0.75 17194

weighted avg 0.90 0.89 0.89 17194

average eval\_loss: 0.4209086316583548, accuracy: 0.890834011864604

training 5 epoch...

1201.3059763908386 seconds used......

5 training finished! average train loss: 0.10250576718384673

total clf loss: 143.45423081633635 total adv loss: 72.7272851197049

evaluating...

Confusion matrix, without normalization

[[ 438 312 272]

[ 383 4801 223]

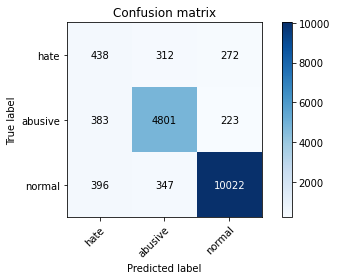
[ 396 347 10022]]

Normalized confusion matrix

[[0.42857143 0.30528376 0.26614481]

[0.0708341 0.88792306 0.04124283]

[0.03678588 0.03223409 0.93098003]]





precision recall f1-score support

hate 0.36 0.43 0.39 1022

abusive 0.88 0.89 0.88 5407

normal 0.95 0.93 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: 0.4598848000715682, accuracy: 0.8875770617657323

2 run...

training 1 epoch...

1200.8452394008636 seconds used......

1 training finished! average train loss: 0.375866326248093

total clf loss: 529.5352239906788 total adv loss: 337.0711100716144

evaluating...

Confusion matrix, without normalization

[[ 455 268 299]

[ 291 4847 269]

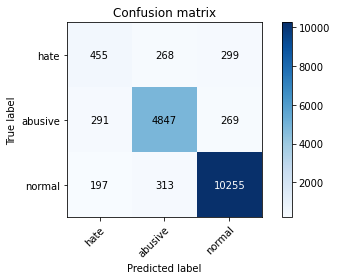
[ 197 313 10255]]

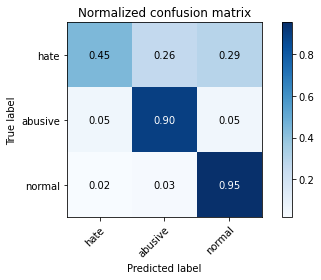
Normalized confusion matrix

[[0.44520548 0.26223092 0.2925636 ]

[0.05381912 0.89643055 0.04975032]

[0.01830005 0.02907571 0.95262425]]





precision recall f1-score support

hate 0.48 0.45 0.46 1022

abusive 0.89 0.90 0.89 5407

normal 0.95 0.95 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: 0.29053955534689646, accuracy: 0.9047923694311969

training 2 epoch...

1201.7150826454163 seconds used......

2 training finished! average train loss: 0.29108949545962065

total clf loss: 409.5649344958365 total adv loss: 250.37726004654542

evaluating...

Confusion matrix, without normalization

[[ 410 342 270]

[ 238 5001 168]

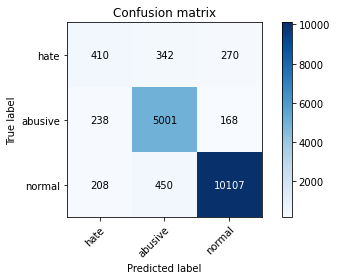
[ 208 450 10107]]

Normalized confusion matrix

[[0.40117417 0.33463796 0.26418787]

[0.04401701 0.92491215 0.03107083]

[0.01932188 0.04180214 0.93887599]]





precision recall f1-score support

hate 0.48 0.40 0.44 1022

abusive 0.86 0.92 0.89 5407

normal 0.96 0.94 0.95 10765

accuracy 0.90 17194

macro avg 0.77 0.75 0.76 17194

weighted avg 0.90 0.90 0.90 17194

average eval\_loss: 0.2905204432648282, accuracy: 0.9025241363266255

training 3 epoch...

1202.0716333389282 seconds used......

3 training finished! average train loss: 0.2153305468037506

total clf loss: 302.3392226230353 total adv loss: 172.56713317986578

evaluating...

Confusion matrix, without normalization

[[ 478 304 240]

[ 338 4851 218]

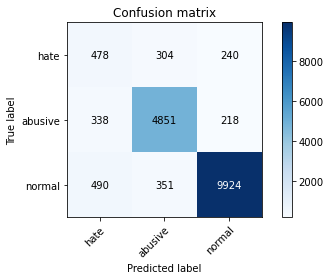
[ 490 351 9924]]

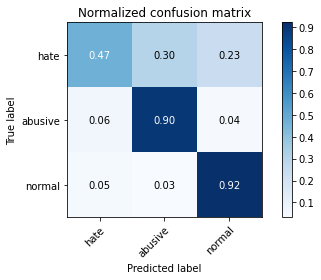
Normalized confusion matrix

[[0.46771037 0.29745597 0.23483366]

[0.06251156 0.89717033 0.04031811]

[0.04551788 0.03260567 0.92187645]]





precision recall f1-score support

hate 0.37 0.47 0.41 1022

abusive 0.88 0.90 0.89 5407

normal 0.96 0.92 0.94 10765

accuracy 0.89 17194

macro avg 0.73 0.76 0.75 17194

weighted avg 0.90 0.89 0.89 17194

average eval\_loss: 0.346929975867618, accuracy: 0.8871117831801791

training 4 epoch...

1201.3751876354218 seconds used......

4 training finished! average train loss: 0.14622998062372372

total clf loss: 204.8427385627292 total adv loss: 107.70090047264239

evaluating...

Confusion matrix, without normalization

[[ 572 238 212]

[ 610 4622 175]

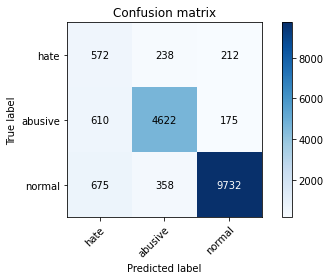
[ 675 358 9732]]

Normalized confusion matrix

[[0.55968689 0.23287671 0.2074364 ]

[0.11281672 0.85481783 0.03236545]

[0.0627032 0.03325592 0.90404087]]





precision recall f1-score support

hate 0.31 0.56 0.40 1022

abusive 0.89 0.85 0.87 5407

normal 0.96 0.90 0.93 10765

accuracy 0.87 17194

macro avg 0.72 0.77 0.73 17194

weighted avg 0.90 0.87 0.88 17194

average eval\_loss: 0.41253764285857597, accuracy: 0.8680935209956961

training 5 epoch...

1201.8419897556305 seconds used......

5 training finished! average train loss: 0.10026505826668584

total clf loss: 140.37861318467185 total adv loss: 72.34147190768272

evaluating...

Confusion matrix, without normalization

[[ 476 299 247]

[ 413 4787 207]

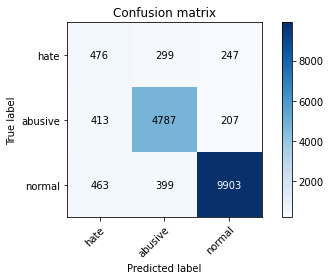
[ 463 399 9903]]

Normalized confusion matrix

[[0.46575342 0.2925636 0.24168297]

[0.07638247 0.88533383 0.03828371]

[0.04300975 0.03706456 0.91992569]]





precision recall f1-score support

hate 0.35 0.47 0.40 1022

abusive 0.87 0.89 0.88 5407

normal 0.96 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: 0.4973269573207078, accuracy: 0.8820518785622892

1 run...

Downloading: 100%

411/411 [00:00<00:00, 12.8kB/s]

Downloading: 100%

263M/263M [00:11<00:00, 23.2MB/s]

training 1 epoch...

1199.6058053970337 seconds used......

1 training finished! average train loss: 0.37631081348377937

total clf loss: 530.1928572840989 total adv loss: 338.09815437346697

evaluating...

Confusion matrix, without normalization

[[ 291 387 344]

[ 116 4966 325]

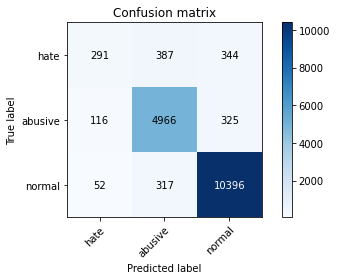
[ 52 317 10396]]

Normalized confusion matrix

[[0.28473581 0.37866928 0.33659491]

[0.02145367 0.91843906 0.06010727]

[0.00483047 0.02944728 0.96572225]]





precision recall f1-score support

hate 0.63 0.28 0.39 1022

abusive 0.88 0.92 0.90 5407

normal 0.94 0.97 0.95 10765

accuracy 0.91 17194

macro avg 0.82 0.72 0.75 17194

weighted avg 0.90 0.91 0.90 17194

average eval\_loss: 0.27732923895033984, accuracy: 0.9103757124578341

training 2 epoch...

1199.22492313385 seconds used......

2 training finished! average train loss: 0.2867305718042787

total clf loss: 403.41510616615415 total adv loss: 246.29213154502213

evaluating...

Confusion matrix, without normalization

[[ 553 221 248]

[ 509 4739 159]

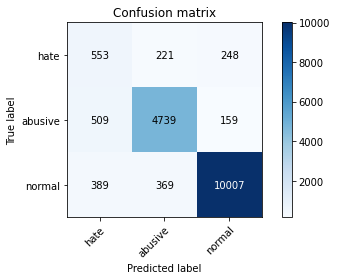
[ 389 369 10007]]

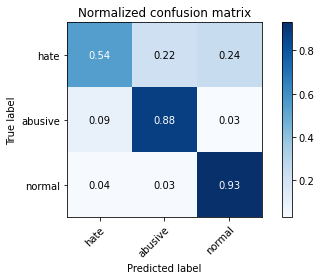
Normalized confusion matrix

[[0.54109589 0.21624266 0.24266145]

[0.09413723 0.87645645 0.02940633]

[0.03613562 0.03427775 0.92958662]]





precision recall f1-score support

hate 0.38 0.54 0.45 1022

abusive 0.89 0.88 0.88 5407

normal 0.96 0.93 0.94 10765

accuracy 0.89 17194

macro avg 0.74 0.78 0.76 17194

weighted avg 0.90 0.89 0.90 17194

average eval\_loss: 0.3242852381408908, accuracy: 0.8897871350471095

training 3 epoch...

1199.6468513011932 seconds used......

3 training finished! average train loss: 0.21336362651710417

total clf loss: 299.59777615591884 total adv loss: 171.3957875703927

evaluating...

Confusion matrix, without normalization

[[ 450 285 287]

[ 294 4860 253]

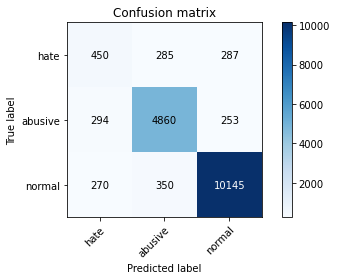
[ 270 350 10145]]

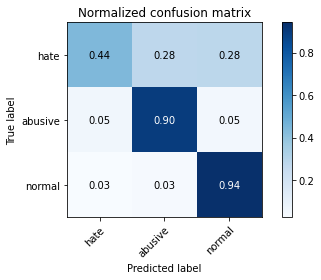
Normalized confusion matrix

[[0.44031311 0.27886497 0.28082192]

[0.05437396 0.89883484 0.0467912 ]

[0.02508128 0.03251277 0.94240595]]





precision recall f1-score support

hate 0.44 0.44 0.44 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 0.76 0.76 17194

weighted avg 0.90 0.90 0.90 17194

average eval\_loss: 0.3239941737701207, accuracy: 0.8988600674653949

training 4 epoch...

1199.6297807693481 seconds used......

4 training finished! average train loss: 0.14179164731173155

total clf loss: 198.5395453092642 total adv loss: 102.71476414101198

evaluating...

Confusion matrix, without normalization

[[ 490 341 191]

[ 305 4958 144]

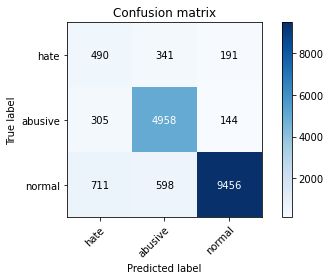
[ 711 598 9456]]

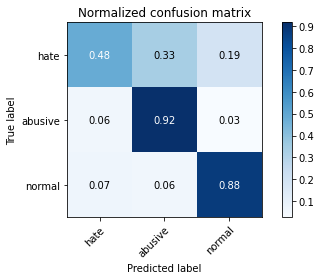
Normalized confusion matrix

[[0.47945205 0.33365949 0.18688845]

[0.05640836 0.9169595 0.02663214]

[0.06604738 0.05555039 0.87840223]]





precision recall f1-score support

hate 0.33 0.48 0.39 1022

abusive 0.84 0.92 0.88 5407

normal 0.97 0.88 0.92 10765

accuracy 0.87 17194

macro avg 0.71 0.76 0.73 17194

weighted avg 0.89 0.87 0.87 17194

average eval\_loss: 0.46815310185341985, accuracy: 0.8668140048854251

training 5 epoch...

1199.5801815986633 seconds used......

5 training finished! average train loss: 0.10056259093350899

total clf loss: 140.81103975366568 total adv loss: 72.87331776024075

evaluating...

Confusion matrix, without normalization

[[ 358 396 268]

[ 211 4963 233]

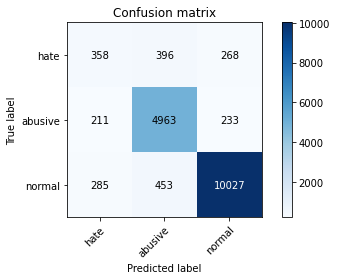
[ 285 453 10027]]

Normalized confusion matrix

[[0.35029354 0.38747554 0.26223092]

[0.03902349 0.91788422 0.04309229]

[0.02647469 0.04208082 0.9314445 ]]





precision recall f1-score support

hate 0.42 0.35 0.38 1022

abusive 0.85 0.92 0.88 5407

normal 0.95 0.93 0.94 10765

accuracy 0.89 17194

macro avg 0.74 0.73 0.74 17194

weighted avg 0.89 0.89 0.89 17194

average eval\_loss: 0.45403092667612055, accuracy: 0.8926369663836222

2 run...

training 1 epoch...

1199.7944486141205 seconds used......

1 training finished! average train loss: 0.37780303717795705

total clf loss: 532.5132341049612 total adv loss: 343.79782863892615

evaluating...

Confusion matrix, without normalization

[[ 325 322 375]

[ 209 4900 298]

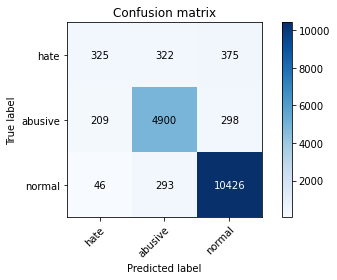
[ 46 293 10426]]

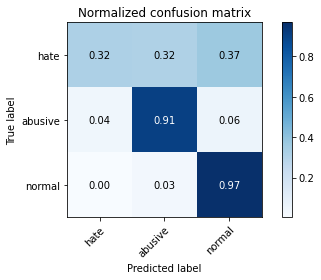
Normalized confusion matrix

[[0.31800391 0.31506849 0.36692759]

[0.0386536 0.90623266 0.05511374]

[0.00427311 0.02721784 0.96850906]]





precision recall f1-score support

hate 0.56 0.32 0.41 1022

abusive 0.89 0.91 0.90 5407

normal 0.94 0.97 0.95 10765

accuracy 0.91 17194

macro avg 0.80 0.73 0.75 17194

weighted avg 0.90 0.91 0.90 17194

average eval\_loss: 0.2884612641947041, accuracy: 0.9102593928114459

training 2 epoch...

1201.320772409439 seconds used......

2 training finished! average train loss: 0.28809938137351243

total clf loss: 405.3861945811659 total adv loss: 248.3727702293545

evaluating...

Confusion matrix, without normalization

[[ 405 331 286]

[ 191 5005 211]

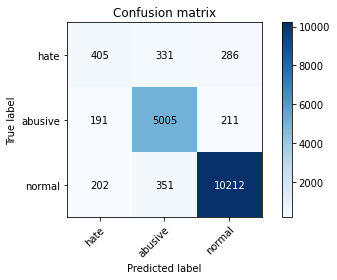
[ 202 351 10212]]

Normalized confusion matrix

[[0.3962818 0.32387476 0.27984344]

[0.03532458 0.92565193 0.03902349]

[0.01876451 0.03260567 0.94862982]]





precision recall f1-score support

hate 0.51 0.40 0.45 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 0.76 0.77 17194

weighted avg 0.90 0.91 0.91 17194

average eval\_loss: 0.2788906448152523, accuracy: 0.9085727579388159

training 3 epoch...

1202.049599647522 seconds used......

3 training finished! average train loss: 0.21408428008466018

total clf loss: 300.6836786447093 total adv loss: 173.45440894830972

evaluating...

Confusion matrix, without normalization

[[ 400 341 281]

[ 224 4974 209]

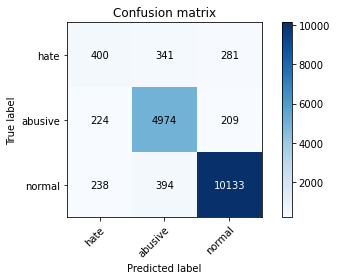
[ 238 394 10133]]

Normalized confusion matrix

[[0.39138943 0.33365949 0.27495108]

[0.04142778 0.91991862 0.0386536 ]

[0.02210869 0.03660009 0.94129122]]





precision recall f1-score support

hate 0.46 0.39 0.42 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 0.76 17194

weighted avg 0.90 0.90 0.90 17194

average eval\_loss: 0.3208897936287834, accuracy: 0.90188437827149

training 4 epoch...

1201.1006848812103 seconds used......

4 training finished! average train loss: 0.14382653422496633

total clf loss: 201.6225863364525 total adv loss: 108.86387378373183

evaluating...

Confusion matrix, without normalization

[[ 392 352 278]

[ 282 4885 240]

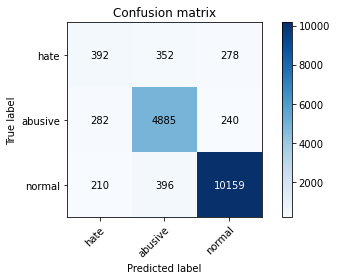
[ 210 396 10159]]

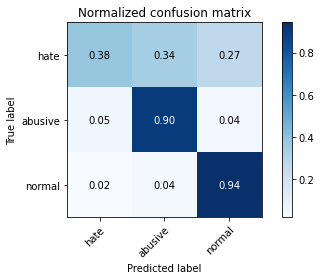
Normalized confusion matrix

[[0.38356164 0.3444227 0.27201566]

[0.05215461 0.90345848 0.04438691]

[0.01950766 0.03678588 0.94370646]]





precision recall f1-score support

hate 0.44 0.38 0.41 1022

abusive 0.87 0.90 0.88 5407

normal 0.95 0.94 0.95 10765

accuracy 0.90 17194

macro avg 0.75 0.74 0.75 17194

weighted avg 0.89 0.90 0.90 17194

average eval\_loss: 0.367154795876271, accuracy: 0.8977550308247063

training 5 epoch...

1200.8226389884949 seconds used......

5 training finished! average train loss: 0.09645772598311032

total clf loss: 134.9898595912964 total adv loss: 68.4304261225625

evaluating...

Confusion matrix, without normalization

[[ 462 283 277]

[ 382 4746 279]

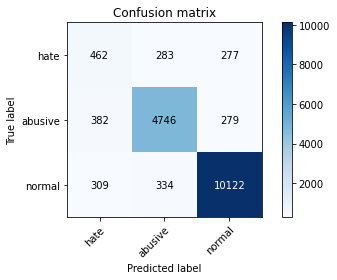
[ 309 334 10122]]

Normalized confusion matrix

[[0.45205479 0.27690802 0.27103718]

[0.07064916 0.87775106 0.05159978]

[0.02870413 0.03102647 0.94026939]]





precision recall f1-score support

hate 0.40 0.45 0.42 1022

abusive 0.88 0.88 0.88 5407

normal 0.95 0.94 0.94 10765

accuracy 0.89 17194

macro avg 0.74 0.76 0.75 17194

weighted avg 0.90 0.89 0.89 17194

average eval\_loss: 0.5038691873136951, accuracy: 0.8915900895661277

1 run

Downloading: 100%

411/411 [00:00<00:00, 14.6kB/s]

Downloading: 100%

263M/263M [00:12<00:00, 21.6MB/s]

training 1 epoch...

1208.5836865901947 seconds used......

1 training finished! average train loss: 0.3728181115731247

total clf loss: 525.0238652229309 total adv loss: 329.99923962540925

evaluating...

Confusion matrix, without normalization

[[ 443 293 286]

[ 250 4927 230]

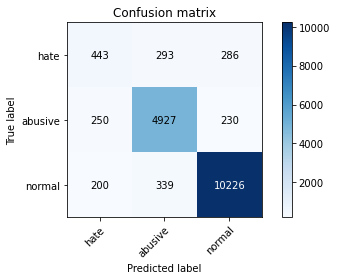
[ 200 339 10226]]

Normalized confusion matrix

[[0.4334638 0.28669276 0.27984344]

[0.04623636 0.91122619 0.04253745]

[0.01857873 0.03149094 0.94993033]]





precision recall f1-score support

hate 0.50 0.43 0.46 1022

abusive 0.89 0.91 0.90 5407

normal 0.95 0.95 0.95 10765

accuracy 0.91 17194

macro avg 0.78 0.76 0.77 17194

weighted avg 0.90 0.91 0.91 17194

average eval\_loss: 0.2946770234943129, accuracy: 0.9070606025357683

training 2 epoch...

1208.1069114208221 seconds used......

2 training finished! average train loss: 0.2902099127141087

total clf loss: 408.2346139512956 total adv loss: 247.76585547812283

evaluating...

Confusion matrix, without normalization

[[ 518 283 221]

[ 331 4894 182]

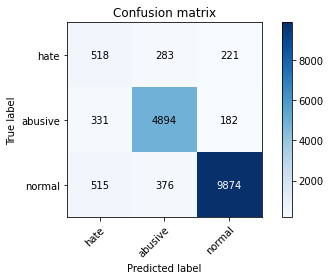
[ 515 376 9874]]

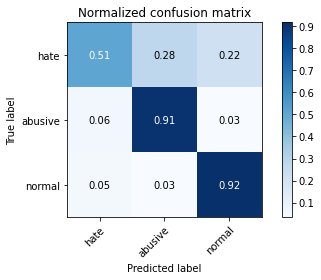
Normalized confusion matrix

[[0.50684932 0.27690802 0.21624266]

[0.06121694 0.90512299 0.03366007]

[0.04784022 0.03492801 0.91723177]]





precision recall f1-score support

hate 0.38 0.51 0.43 1022

abusive 0.88 0.91 0.89 5407

normal 0.96 0.92 0.94 10765

accuracy 0.89 17194

macro avg 0.74 0.78 0.76 17194

weighted avg 0.90 0.89 0.89 17194

average eval\_loss: 0.31928790400541107, accuracy: 0.8890310573455856

training 3 epoch...

1206.6570029258728 seconds used......

3 training finished! average train loss: 0.2157222470896772

total clf loss: 302.75444056931883 total adv loss: 170.1859057676047

evaluating...

Confusion matrix, without normalization

[[ 469 308 245]

[ 303 4894 210]

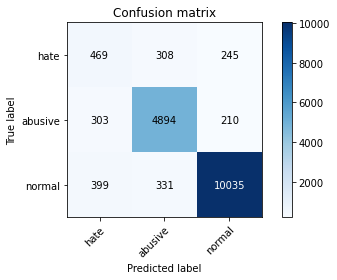
[ 399 331 10035]]

Normalized confusion matrix

[[0.45890411 0.30136986 0.23972603]

[0.05603847 0.90512299 0.03883854]

[0.03706456 0.03074779 0.93218765]]





precision recall f1-score support

hate 0.40 0.46 0.43 1022

abusive 0.88 0.91 0.89 5407

normal 0.96 0.93 0.94 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: 0.3241470905518982, accuracy: 0.8955449575433291

training 4 epoch...

1206.404547214508 seconds used......

4 training finished! average train loss: 0.1483117162800992

total clf loss: 207.832238759147 total adv loss: 110.7011594719952

evaluating...

Confusion matrix, without normalization

[[ 468 322 232]

[ 366 4853 188]

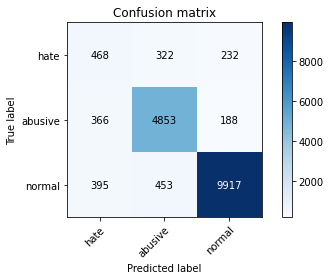
[ 395 453 9917]]

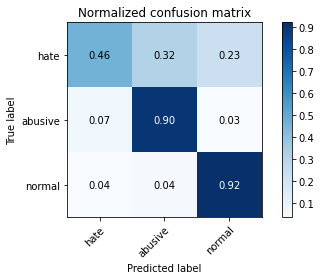
Normalized confusion matrix

[[0.45792564 0.31506849 0.22700587]

[0.06769003 0.89754023 0.03476974]

[0.03669299 0.04208082 0.9212262 ]]





precision recall f1-score support

hate 0.38 0.46 0.42 1022

abusive 0.86 0.90 0.88 5407

normal 0.96 0.92 0.94 10765

accuracy 0.89 17194

macro avg 0.73 0.76 0.75 17194

weighted avg 0.89 0.89 0.89 17194

average eval\_loss: 0.40860073789304424, accuracy: 0.886239385832267

training 5 epoch...

1206.7188460826874 seconds used......

5 training finished! average train loss: 0.10098288399810887

total clf loss: 141.4283872564556 total adv loss: 73.75466699343815

evaluating...

Confusion matrix, without normalization

[[ 456 271 295]

[ 434 4666 307]

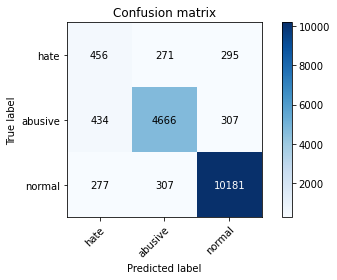
[ 277 307 10181]]

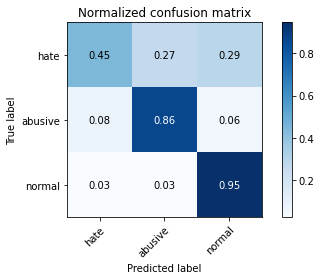
Normalized confusion matrix

[[0.44618395 0.26516634 0.28864971]

[0.08026632 0.86295543 0.05677825]

[0.02573154 0.02851835 0.94575012]]





precision recall f1-score support

hate 0.39 0.45 0.42 1022

abusive 0.89 0.86 0.88 5407

normal 0.94 0.95 0.94 10765

accuracy 0.89 17194

macro avg 0.74 0.75 0.75 17194

weighted avg 0.89 0.89 0.89 17194

average eval\_loss: 0.47825823839277376, accuracy: 0.890019774339886

2 run...

training 1 epoch...

1206.7127392292023 seconds used......

1 training finished! average train loss: 0.37011653010233625

total clf loss: 521.4184833616018 total adv loss: 331.5907163377851

evaluating...

Confusion matrix, without normalization

[[ 299 408 315]

[ 106 5098 203]

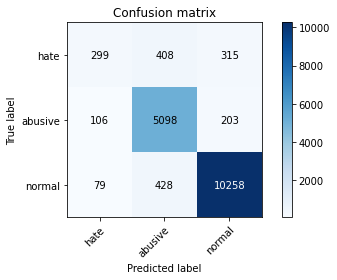
[ 79 428 10258]]

Normalized confusion matrix

[[0.2925636 0.39921722 0.30821918]

[0.01960422 0.94285186 0.03754392]

[0.0073386 0.03975848 0.95290293]]





precision recall f1-score support

hate 0.62 0.29 0.40 1022

abusive 0.86 0.94 0.90 5407

normal 0.95 0.95 0.95 10765

accuracy 0.91 17194

macro avg 0.81 0.73 0.75 17194

weighted avg 0.90 0.91 0.90 17194

average eval\_loss: 0.2773753330024869, accuracy: 0.9104920321042224

training 2 epoch...

1207.2641322612762 seconds used......

2 training finished! average train loss: 0.2886053172754031

total clf loss: 406.01992212608457 total adv loss: 247.24538716021925

evaluating...

Confusion matrix, without normalization

[[ 460 287 275]

[ 289 4884 234]

[ 419 333 10013]]

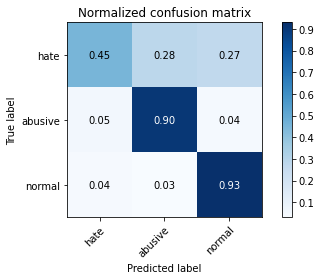
Normalized confusion matrix

[[0.45009785 0.28082192 0.26908023]

[0.05344923 0.90327353 0.04327723]

[0.03892243 0.03093358 0.93014399]]





precision recall f1-score support

hate 0.39 0.45 0.42 1022

abusive 0.89 0.90 0.90 5407

normal 0.95 0.93 0.94 10765

accuracy 0.89 17194

macro avg 0.74 0.76 0.75 17194

weighted avg 0.90 0.89 0.90 17194

average eval\_loss: 0.31163503146015625, accuracy: 0.8931604047923695

training 3 epoch...

1206.8015983104706 seconds used......

3 training finished! average train loss: 0.21708331310771994

total clf loss: 304.89992313459516 total adv loss: 175.96568346954882

evaluating...

Confusion matrix, without normalization

[[ 426 348 248]

[ 223 5006 178]

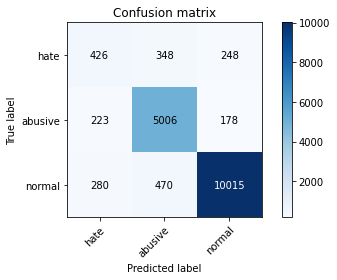
[ 280 470 10015]]

Normalized confusion matrix

[[0.41682975 0.34050881 0.24266145]

[0.04124283 0.92583688 0.03292029]

[0.02601022 0.04366001 0.93032977]]





precision recall f1-score support

hate 0.46 0.42 0.44 1022

abusive 0.86 0.93 0.89 5407

normal 0.96 0.93 0.94 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: 0.33186912612534714, accuracy: 0.8983947888798418

training 4 epoch...

1206.94109416008 seconds used......

4 training finished! average train loss: 0.14561162530156085

total clf loss: 203.89494610286783 total adv loss: 105.6137865439523

evaluating...

Confusion matrix, without normalization

[[ 495 268 259]

[ 454 4729 224]

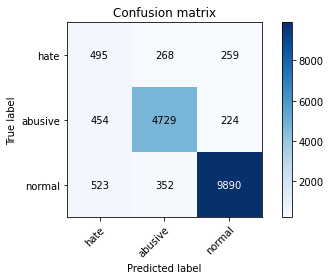
[ 523 352 9890]]

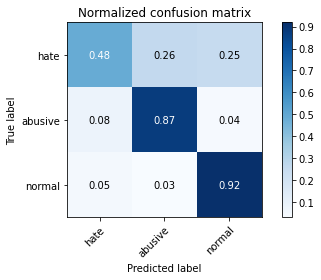
Normalized confusion matrix

[[0.48434442 0.26223092 0.25342466]

[0.08396523 0.87460699 0.04142778]

[0.04858337 0.03269856 0.91871807]]





precision recall f1-score support

hate 0.34 0.48 0.40 1022

abusive 0.88 0.87 0.88 5407

normal 0.95 0.92 0.94 10765

accuracy 0.88 17194

macro avg 0.72 0.76 0.74 17194

weighted avg 0.89 0.88 0.89 17194

average eval\_loss: 0.39929589585314484, accuracy: 0.8790275677561941

training 5 epoch...

1207.2279770374298 seconds used......

5 training finished! average train loss: 0.10081325214613254

total clf loss: 141.01106287981384 total adv loss: 70.03573842043988

evaluating...

Confusion matrix, without normalization

[[ 487 289 246]

[ 411 4722 274]

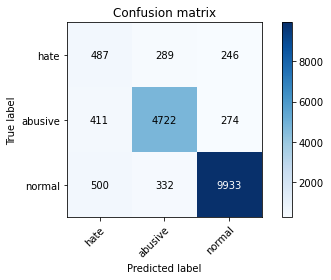
[ 500 332 9933]]

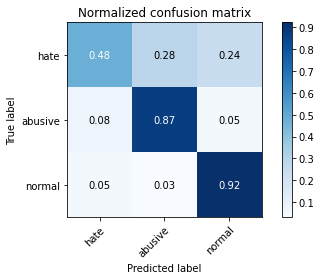
Normalized confusion matrix

[[0.47651663 0.28277886 0.2407045 ]

[0.07601258 0.87331237 0.05067505]

[0.04644682 0.03084069 0.92271249]]





precision recall f1-score support

hate 0.35 0.48 0.40 1022

abusive 0.88 0.87 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: 0.4844064868699629, accuracy: 0.8806560428056299